

On the campaign trail

The use and abuse of science is emerging as an issue in the US presidential election. Researchers should seize an opportunity to make their voices heard, whatever their political persuasion.

Last week's Democratic National Convention in Boston was all about projecting an image of presidential candidate John Kerry that was designed to resonate with undecided voters. We were given Kerry the family man and Kerry the war hero — “reporting for duty” as the next commander-in-chief. But another character also made a cameo appearance: Kerry the friend of science.

Among the star speakers at the convention was Ron Reagan, who argued passionately for research on human embryonic stem cells to help those — like his father Ronald, the late Republican president — who are cruelly struck down by degenerative diseases. Kerry built on that theme. And in her speech, his wife, Teresa Heinz Kerry, lauded NASA's Hubble Space Telescope and its Cassini mission to Saturn. Democratic strategists have evidently figured that there are votes to be gained in supporting science, and that this is an area of policy in which President George W. Bush is perceived to be weak.

The Bush administration has been heavily criticized in scientific quarters. In the spring, the left-leaning Union of Concerned Scientists (UCS) received favourable media coverage when it slammed the Bush administration for stacking advisory committees with its supporters and ignoring advice that runs counter to its ideology. The UCS report was accompanied by a statement from 62 prominent researchers calling on Bush to put a stop to the “politicization of science”.

Many Republican politicians saw the statement — occurring as it did, in an election year — as a political act in itself. And in June, 48 Nobel laureates took things a step further by publicly endorsing the Democratic contender. “John Kerry will restore science to its appropriate place in government and bring it back into the White House.

He is the clear choice for America's next President,” they wrote. Now, like-minded scientists are planning to form a new group, Scientists and Engineers for Kerry, in the hope of swinging the election in favour of the Democratic hopeful (see page 595).

In the current polarized political climate, it is hardly surprising that some scientists should swing behind Kerry in this way — the research community traditionally votes overwhelmingly Democratic. But the emergence of science as an election issue presents opportunities that cut across the lines of partisan party politics.

Most researchers are not overtly political animals. They are more comfortable arguing for a policy that they believe to be based on sound science, rather than campaigning for one party or another. For the time being, Kerry is making the running on scientific issues. But if his advisers are correct in their judgement that the public is receptive to a pro-science message, it won't be long before the Bush campaign is seeking to build bridges with the scientific community.

Scientists, regardless of their political affiliation, should take advantage of this course of events. Whoever wins or loses, scientists have an opportunity to raise the profile of research nationwide. And they may cause whoever triumphs in November to think more deeply about how they will use science in their administration.

Above all, researchers should remember that the last election was decided by just 537 votes in Florida — less than one-eighth of the number of scientists who have so far signed the statement backing the UCS report. Both Bush and Kerry should be reminded that, in a close-run poll, science-friendly policies could help swing the balance between defeat and victory. ■

Weak at the centre

Germany's obsessive federalism has become an anachronism that is holding back the nation's science.

To those outside Germany, it seems inconceivable that a failed attempt to discard an archaic academic qualification could displace the tragedies of Iraq and Sudan from the news headlines. But it happened last week, after Germany's constitutional court ruled that the central government had no right to tell the country's states, or *Länder*, how to run their universities (see page 599).

The ruling has a wider political significance. In 1949, Germany's postwar constitution handed primary responsibility for many areas of policy to the *Länder*, to prevent anyone from emulating Adolf Hitler's rise to supreme power. Today, however, this decentralization has become a handicap to Germany's competitiveness.

The current débâcle surrounds attempts to replace the outmoded *Habilitation*, a post-PhD qualification traditionally required for progress in German academia. The mess arose from a laudable attempt by federal research and education minister Edelgard Bulmahn to modernize the German academic system by establishing ‘junior professorships’ as an alternative. But three conservative *Länder* objected to Bulmahn's intervention in the affairs of their universities, and took the case to the constitutional court.

What's most depressing is that the three *Länder* have nothing

against the idea of providing independent positions for talented young scientists — they were just defending turf. And this isn't an isolated example: last month, *Länder* governments delayed, perhaps for ever, the introduction of a €1.9-billion (US\$2.3-billion) windfall to upgrade some universities to ‘élite’ status (see *Nature* **430**, 283; 2004). Before the money is handed out, they wanted the relevant responsibilities of the central and *Länder* governments to be clarified by a high-level ‘federalism committee’, which was set up last autumn.

Conservative *Länder* politicians — led by Edmund Stoiber, prime minister of Bavaria and co-chair of the federalism committee — are now pushing to scrap a 1976 law that shares responsibility for science and higher education between the *Länder* and the central government. They want to devolve more power to themselves.

This would be a retrograde move. If individual *Länder* draw up their own policies on academic career development, moving between universities in Germany could become as complicated as going abroad. The German academic system could degenerate into a series of inward-looking institutions, none of them able to compete on the international stage. What German academia needs is more centralized strategic thinking, not less. ■





Rising tide
Flash floods in Asia prompt calls for land reform
p596



Active to the end
Colleagues pay tribute to Francis Crick
p597



Mini menace?
Threat posed by nanoparticles not to be sniffed at
p599



Pooled resources
Amateur divers join initiative to assess the state of the sea
p601

Nobel laureates spearhead effort to put Kerry in the White House

Geoff Brumfiel and Emma Marris, Washington

A group of high-profile researchers is to hit the campaign trail on behalf of Democratic presidential candidate John Kerry.

Planners hope that the group, known as Scientists and Engineers for Kerry, will turn researchers' anger over the Bush administration's treatment of science into a major political force. In the run-up to the election, they will attempt to recruit scientists across the country to deliver speeches on Kerry's behalf and to campaign for his election.

"It's very hard to anticipate, but I think we're going to get a very strong response," says Henry Kelly, president of the Federation of American Scientists in Washington, who is working independently from that organization to spearhead the group. Kelly, who helped put together a recent letter from 48 Nobel prizewinners endorsing the Kerry campaign (see *Nature* 430, 4; 2004), believes that the group of researchers will be far larger than any ever brought together on behalf of a presidential candidate.

Kelly has so far secured three Nobel laureates from across the scientific disciplines to head the group. Harold Varmus, a geneticist and former director of the National Institutes of Health, has agreed to co-chair the committee with Mario Molina, an atmospheric chemist at the Massachusetts Institute of Technology, and Burton Richter, a physicist and director emeritus of the Stanford Linear Accelerator Center in California. Others may join them.

"I have never trotted out my Nobel prize for a political campaign," says Richter. But he says recent allegations that the Bush administration has manipulated committees and worked to misrepresent and suppress academic findings have stirred him into action. "Under Bush, science is being distorted for political ends," he says.

This is not the first time that researchers have lined up behind a presidential candidate. Most recently, a 1992 group calling itself Scientists and Engineers for Clinton/Gore brought together more than

60 researchers, including 12 Nobel laureates, to endorse the then presidential hopeful. But that group met with only modest success, organizing relatively few events and meeting only once with a high-ranking member of the campaign — Al Gore, who was then the vice-presidential candidate.

Lewis Branscomb, a professor of public policy at Harvard University who has advised both Democratic and Republican administrations on scientific issues, says he is sceptical as to whether the new group will have an impact on the election this November. "I doubt it will amount to much," he says. Despite signing a statement by the Union of Concerned Scientists denouncing the Bush administration's treatment of science, Branscomb says he is unlikely to join the Kerry group.

Impact factor

David Leiter, who spent six years as Kerry's chief of staff and is currently helping in his campaign, says he is hopeful that the group will have a real impact. "The Bush administration is disinvested in science and R&D and who better to bring that to the public's attention than scientists and engineers," he says.

Already, science has taken an unusually high profile in the Kerry campaign. Kerry's wife, Teresa Heinz Kerry, mentioned the Hubble Space Telescope and the Cassini mission to Saturn in her speech at the Democratic National Convention last week.

And stem-cell research has emerged as a particularly hot issue. In his speech officially accepting his nomination as the Democrat candidate last week, Kerry asked: "What if we have a president who believes in science, so we can unleash the wonders of discovery like stem-cell research to treat illness and save millions of lives?" Ron Reagan, the son of two-term Republican president Ronald Reagan, spoke at length during the convention about the promise of stem cells and criticized Bush's policies, which limit the number of cell lines available for federally funded research.

No firm date has been set for the launch of the scientists' group, but Kelly is optimistic that it will be up and running this month. ■



US researchers hope to make science a key issue in John Kerry's campaign to become president.

Lab closure sparks fears that US prion research is waning

Helen Pearson, New York

A major lab dedicated to studying prions closed down last week, causing some to worry about the future of research on these proteins in the United States. But at the same time, another US lab revealed findings emphasizing the importance of prions in brain-wasting diseases.

The Laboratory of Central Nervous System Studies at the National Institute of Neurological Disorders and Stroke (NINDS) in Bethesda, Maryland, has long conducted research on prions. These are the infectious agents thought to cause conditions such as mad cow disease and its human version, variant Creutzfeldt-Jakob disease.

Work at NINDS has been winding down for some years and came to a halt on 30 July with the retirement of its medical director Paul Brown. The closure leaves the National Institutes of Health (NIH) with just one major in-house prion research facility, at Rocky Mountain Laboratories in Hamilton, Montana.

"The idea of losing the whole NINDS facility is devastating," says virologist Dean Cliver of the University of California, Davis, who has advised government agencies on prion-research strategy. The closure reflects the lack of interest in prion research in the United States, he adds.

Bruce Chesebro, who heads the prion lab at Rocky Mountain, says that NIH labs with secure, long-term funding are important for prion research because animals take months or years to develop symptoms. Last November, the Institute of Medicine also recommended that the NIH should strengthen its in-house prion research.

A handful of other US labs study prions, including Stanley Prusiner's facility at the University of California, San Francisco. Prusiner won a Nobel prize in 1997 for his work in identifying prions as a disease-causing agent. This week, he and his group announced that they had created a prion protein in bacteria and shown that it caused disease when injected into the brains of mice (G. Legname *et al.* *Science* 305, 673–676; 2004), strengthening the hypothesis that prions cause disease without the help of genetic material.

US scientists acknowledge the quality of such work, but say they need more labs to keep up with the pace of research in Europe. The NINDS is considering recruiting another lab in the field, says Eugene Major, acting director of the institute's neuroscience programme. ■



Muddy waters: scientists are striving to clarify the issues as Bangladesh is hit by devastating floods.

Reform of land use urged as floodwaters rise across Asia

David Cyranoski, Tokyo

Storms and floods ravaging Asia this summer have claimed the lives of more than 1,500 people and left tens of millions facing food shortages and water-borne diseases. While relief agencies are busy sending aid, scientists are working out how to prevent such disasters from happening again.

Several factors have conspired to make the floods in the northeastern region of Bangladesh the worst in recorded history, says Atiq Rahman, head of the Bangladesh Centre for Advanced Studies in Dhaka. Four flash floods came in quick succession, he says, and unusually high tides have slowed the water's discharge into the sea. Little can be done about such conditions, but we can act to lessen their impact, he says.

The Dhaka-based Flood Forecasting and Warning Center, for example, can give three to seven days' notice of floods. This time could be used to focus relief efforts, says Rahman, allowing emergency supplies to get to the right areas of the country. "We knew which areas would be affected most. There was enough warning," he says, adding that better political organization was all that was needed.

Changes to land use could also have a big effect, says Vaidyanatha Subramanian, a hydrogeochemist at Jawaharlal Nehru University in New Delhi. Canals in Bangladesh have been squeezed by land development, he says, increasing the chance that they will burst from their confines. River banks have been eroded by unregulated mining for sand and by bathing at religious festivals. Subramanian is now collecting data about the timing, strength and duration of rains and the resulting river flow, to work out how much of the flooding is caused by nature rather than humans.

In Taiwan, where typhoon Mindulle caused particularly bad floods in early July, scientific experts have visited disaster sites to work out what went wrong. They note that much of the problem was beyond the government's control: the rains were particularly heavy and the 1999 Chi-Chi earthquake left loose soil that was easily swept away. But they also identified the farming of tea and betelnuts on steep slopes as a factor in excessive soil erosion. The government's subsequent decision to shut down three farming areas has provoked tension with local people.

Bangladesh must also cope with social tension over its land use. Some of the richer areas near hard-hit Dhaka were kept dry by massive embankments. Reports say police used rubber bullets to break up protests where angry citizens tried to break open sluice gates in attempts to open natural drainage routes for the water.

Some of Bangladesh's woes are being addressed by a National Water Management Plan, approved by the National Water Research Council this March. The plan aims to decentralize decision-making about flood management, and create a central organization for sharing information. But the projects, to be implemented over 25 years, have had little time to take effect, says Rahman.

As of 2 August, water levels had dropped to normal throughout most of Bangladesh. But that is little comfort, says Rahman. More rain, expected in mid-August, will create havoc in the saturated area. "The next round will be worse," he says. ■

For more news and analysis go to
news @ nature.com
www.nature.com/news

From DNA to consciousness — Crick's legacy

Jonathan Knight

Francis Crick, widely regarded as one of the twentieth century's most significant figures in biology, was a scientist to the end. In his hospital bed, just hours before he died on 28 July following a prolonged battle with colon cancer, he was working on a manuscript.

It was a theoretical discussion of the claustrum, a little-studied region of the brain that might play an important role in human consciousness, the object of Crick's academic efforts for nearly 30 years. His unending dedication to his work came as no surprise to friends and colleagues who spoke about him this week.

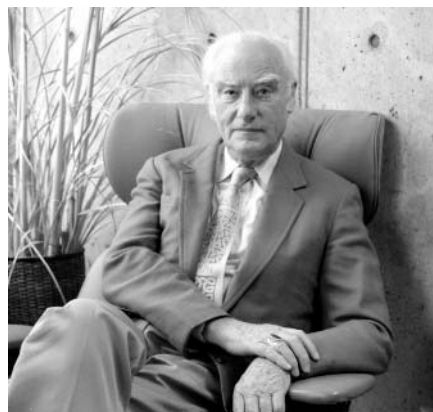
"He was the living incarnation of what it is to be a scholar," says Christof Koch, a computational neuroscientist at the California Institute of Technology in Pasadena and a close colleague of Crick's for 16 years. "He was always willing to revise his own views in light of the actions of a Universe that never ceased to astonish him."

Although Crick's name will forever be connected with the discovery of the double-helix structure of DNA, which he published in 1953 with James Watson (*Nature* 171,

737–738; 1953), his subsequent contributions to molecular biology and neuroscience have also had a profound impact.

In the 1950s, he helped to prove that sequences of three bases in DNA code for particular amino acids, and described how these could be used to make proteins. Later he was instrumental in decrypting which triplets of bases coded for which amino acids.

By the mid-1970s, Crick's interest in



Francis Crick's unending dedication to science saw him working to the very end.

molecular biology had begun to wane. He felt that the important problems had been solved, and that only details remained to be worked out. "He didn't want to work on small problems," recalls Alexander Rich, a biologist at the Massachusetts Institute of Technology in Cambridge who solved the structure of one of the body's main structural proteins — collagen — with Crick in 1955 (A. Rich and F.H.C. Crick *Nature* 176, 915–916; 1955).

In 1976, Crick moved to the Salk Institute for Biological Sciences in La Jolla, California, where he would devote the rest of his life to understanding the biological basis of consciousness — a problem many considered too difficult to tackle. "Because of his stature, he made brain science respectable," says Leslie Orgel, a colleague of Crick's at Salk.

The paper Crick was finishing when he died may influence the direction of Salk's Crick–Jacobs Center for Computational and Theoretical Biology, established earlier this year, says president Richard Murphy. "He thought the claustrum would be a good test system for the centre," Murphy says. The centre will focus on the genes, proteins and neural networks that make the brain function. ■

► www.nature.com/nature/focus/crick

M. LIEBERMAN/SALK

'Militant' animal activists trigger British law change

Jim Giles, London

"Exposed: mastermind of training camp for militants" read a front-page headline in the *London Evening Standard* last Wednesday. But this was not a story about al-Qaeda or political terrorists — the militants in question were animal-rights campaigners.

Animal-rights activists have led sometimes violent protests against biomedical research for more than a decade, particularly in Europe. But last week saw an upsurge in the attention paid to their campaigns in Britain, with the introduction of new laws against their protests and the announcement of fresh money for animal research.

Legislation unveiled on 30 July makes it illegal to mount protests that cause "harassment, alarm or distress" outside people's homes. Harassing a company's employees will also become an arrestable offence.

Such rules are seen as necessary by many scientists, given the history of animal-rights activism in Britain. Over the past ten years, supporters of animal research have been attacked with baseball bats and

had letter bombs posted to their homes. The *Standard* article claimed to reveal one person who was prepared to train "fanatics" in "unarmed combat" and "infiltration techniques" for future campaigns.

Recently, activists have focused their attention on plans for large, centralized research facilities. In January, one campaign contributed to the collapse of the University of Cambridge's plans for a new primate research centre. Similarly, on 19 July, building contractor Montpellier pulled out of work on an animal-research facility at the University of Oxford after attacks by

activists; the university insists that the project will go ahead.

In the same week as the new legislation, pharmaceutical companies pledged £4 million (US\$7 million) over four years to fund animal research for medicines. The money pledged is small compared with the hundreds of millions spent in Britain each year on animal research, but scientists see it as a welcome gesture of support.

Observers say that both moves reflect a general change in public opinion. "People are increasingly aware of how our health depends on animal research," says Tipu Aziz, a neurosurgeon at the University of Oxford who is one of only a handful of researchers prepared to speak publicly in favour of animal research.

But lobbyists caution that it is not clear how successful the new legislation will be at deterring protests. This is the third set of such UK laws introduced since 2001, but the campaigners have proved adept at switching tactics to avoid prosecution, says Mark Matfield, executive director of the RDS, a London-based organization that lobbies in favour of animal research. "The government is focusing on protestors' tactics," says Matfield. "We need to look at making it an offence to organize these campaigns." ■



Campaigns against large research facilities have escalated.

P. HAYS/REX FEATURES

Review of tenure refusal uncovers conflicts of interest

Rex Dalton, San Diego

The academic rights of an ecologist at the University of California, Berkeley, may have been violated when he was denied tenure last year, according to a report from the academic senate.

Ignacio Chapela was an outspoken critic of Berkeley's controversial academic-industrial partnership with the Swiss agribiotech firm Syngenta, which ended last year (see right). He was also the lead author of a disputed paper in *Nature* in which he asserted that genes from genetically modified crops had flowed into Mexican maize, and had become scattered throughout the genome (D. Quist and I. H. Chapela *Nature* 414, 541–543; 2001). After a

storm of criticism about the paper, *Nature* withdrew its support for the article, but the authors stand by their research.

Against this background, Chapela was denied tenure at Berkeley's College of Natural Resources in November 2003 (see *Nature* 426, 591; 2003). He appealed.

The resulting report, issued on 28 June, claims that Jasper Rine, a geneticist at the university who sat on a key committee reviewing Chapela's tenure, had conflicts of interest. It says that Rine had financial dealings with biotech firms, oversaw the Syngenta agreement and had cited Chapela's *Nature* paper as an example of poor science in one of his classes. Both the dean of Chapela's college and his department chair requested that Rine be taken from the committee four times; but Rine did not excuse himself nor did the committee chair ask him to leave. The report also says there was "unjustifiable" delay in the tenure-review process.

"I am glad the senate is able to rise to the occasion," says Chapela, whose contract has been extended while he appeals. Rine, Berkeley administrators and senate members would not comment on the report, citing its confidentiality. Such committee reports are rarely disclosed.

As the senate continues its inquiry, Chapela is hoping for a second tenure review. He has also filed two claims that may precede a lawsuit. In April, he accused the university of discrimination, saying that he was denied tenure because he is Hispanic. Early last month, he claimed he was victimized by the university for speaking out against the Syngenta deal. ■



Ignacio Chapela: his appeal over being denied tenure is ongoing.

Rex Dalton, San Diego

Large-scale partnerships between industry and universities ought to be avoided completely, according to the author of an external report investigating such a deal at the University of California at Berkeley. Although the arrangement provided cash that energized research, the report says, it created conflicts of interest within the university.

The report was commissioned by the university's academic senate and conducted by Lawrence Busch, an agricultural sociologist at Michigan State University, East Lansing. It looks in detail at a controversial deal struck in 1998 between the university and a Swiss firm now known as Syngenta. The five-year agreement saw Syngenta contribute US\$25 million to facilities and the funding of faculty at the university's department of plant and microbial biology, in exchange for the right to capitalize on their discoveries. The deal ended last year; Syngenta officials say it was abandoned owing to a change in company strategy.

Following their two-year review, Busch and his eight colleagues conclude that the deal did not compromise academic research



Tart reception: those announcing the private-public partnership were rebuffed with a pie.

Biotech funding deal judged to be 'a mistake' for Berkeley

as many scientists had feared. But it did produce institutional conflicts of interest, Busch says, contributing to the denial of tenure to a faculty member who was critical of the deal (see left). It also failed to produce as many patent benefits as originally hoped.

Busch hopes the report will lead to a broader debate over academic-industrial relations at all US universities. "I think it is high time for serious discussions of what the devil we want our universities to be. Berkeley is a good place to start," says Busch. "I don't think you should hermetically insulate a university from the private sector. But this deal was a mistake." Most university-industry projects fund specific researchers or scientific groups rather than an entire academic unit, as the Berkeley pact did, notes Busch.

When the deal in question was initially announced (with Novartis, whose agriculture business became Syngenta), it was so loathed by students and scientists at the university that demonstrators threw a pie at the officials announcing it. There was a widespread perception that the Syngenta deal "compromised the mission of the university," says Busch. The academic senate pushed for and won university funding to conduct a \$225,000 analysis of the deal.

Commissioning an independent report on such an issue is an unusual move, says Sheldon Krinsky, a philosopher at Tufts University, Massachusetts, who studies academic institutions. If it were not for the academic senate's concerns, certain criticisms might never have come to light, he says. Berkeley's administrators conducted an internal review of the deal two years ago; they concluded that the Syngenta pact was a good deal.

The Busch report highlights how the money injected into the university did make a difference: 26 members of the faculty received financial support from Syngenta, with individual totals ranging from \$60,000 to \$200,000. And, during the five-year period, the department made 51 potentially patentable discoveries, 12 of which emerged from Syngenta-funded research. But, out of 20 patents produced, Syngenta is pursuing research on only six, and no licence agreements with the university have been negotiated for any of those.

The university's administrative leaders declined to be interviewed, but the provost Paul Gray issued a statement saying the report is under review. Collaborations are important to the university's mission, the statement says, but "the appropriateness and structure" of such agreements should be examined. Syngenta officials did not respond to *Nature's* request for an interview. ■

Court ruling upsets hopes for career reforms

Alison Abbott, Munich

Germany's young scientists suffered a blow this week when one of their most favoured career paths was declared unconstitutional.

In 2002, Germany introduced the title of 'junior professor' in a bid to reform the higher-education system and give young scientists more independence. But on 27 July the country's highest court found that the federal government had no right to force the system on Germany's 16 individual *Länder*, or states.

The 600 young scientists who have already won these junior positions are expected to keep their jobs, but the creation of future posts is uncertain. The ruling has been met with horror by critics, who say it throws the country's education system back into the dark ages. They say that Germany's outdated academic system is harming its international competitiveness.

"We are very unhappy — this is a real throwback into provincialism," says Julia Fischer, speaker for the Young Academy, which promotes young German scientists, and a group leader at the Max Planck Institute for Evolutionary Anthropology in Leipzig.

Researchers in Germany have long been frustrated by the traditional system for advancement. After achieving a PhD, scien-



Fighting on: Edelgard Bulmahn will work to save the endangered 'junior professor' post.

tists work with an established professor to complete a second thesis and prove their teaching abilities — a process known as *Habilitation*. Many German academics do not gain tenure until their early forties.

The scientific community, including the influential Wissenschaftsrat, or science council, and the DFG, the country's main funding agency, have long called for *Habilitation* to be replaced by a US-style tenure-track system. In 2002, the position of junior professor was created in response.

Although not linked to tenure track, the

position at least allowed German thirty-somethings to prove their academic merit independently. The junior professors received some €60,000 (US\$70,000) to set up their research over a period of up to six years. Federal education minister Edelgard Bulmahn set aside €180 million for the programme, and stated that *Habilitation* would be phased out.

By 2004, nine German states had adopted the federal law into their own rule books. But three conservative states — Bavaria, Saxony and Thuringia — took the issue to court over the principle that the federal government should not interfere with their affairs in education. The court agreed.

It is now unclear what will happen to both the money and the jobs. Some states, including Berlin, have vowed to keep the new career path open. Bulmahn says that she will work to modify the law to try and maintain the junior professorship.

But many, including Fischer, argue that the academic system will not be acceptable until *Habilitation* disappears completely from all 16 states. Even when other career paths exist, most young German scientists, including herself, feel obliged to do a *Habilitation* to keep their options open. "You never know whether you will be judged badly if you don't," she says.

Size matters when it comes to safety, report warns

Jim Giles, London

Manufactured nanoparticles should be treated as dangerous until proven harmless, according to Britain's Royal Society and Royal Academy of Engineering. And that requires an overhaul of safety regulations both in and outside the laboratory, they say.

The two organizations were asked to examine the safety of nanotechnology by the UK government. Their report, published on 29 July, calls for the law to recognize something scientists already acknowledge: substances that are safe in large quantities can be dangerous when reduced to particles on the scale of a billionth of a metre.

"Where particles are concerned, size really does matter," says Ann Dowling, a mechanical engineer at the University of Cambridge and chair of the group that produced the report.

No one yet knows whether any particular nanoparticles can harm human or environmental health. But there are hints that large doses of some particles can be dangerous to animals, and a study soon to be published shows that inhaled carbon nanoparticles can make their way to the

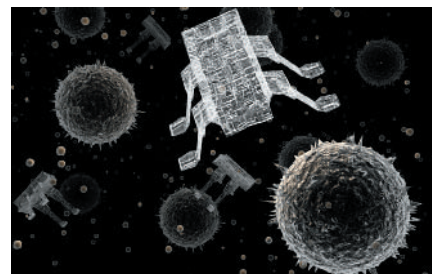
brain (G. Oberdörster *Inhalation Toxicol.* in the press).

Nanoparticles are already present in a range of materials, from engine exhaust fumes to sunscreens. Analysts predict that the use of nano-sized tubes and wires will have a big impact on many future products, from drug delivery systems to next-generation computer chips.

The report says that any use of manufactured nanoparticles should be assessed by dedicated scientific panels, operating alongside the regulatory bodies that currently assess the safety of chemicals. All the data should be made public, the authors add.

They also call for researchers to use fume hoods and other protection when handling the particles. Dowling says that many scientists already take such measures, but adds that some labs are currently unaware of the potential dangers.

Dowling's panel also wants Britain's research councils — the main source of UK grants for basic research — to create a research centre dedicated to the study of environmental and human-health aspects



Nano-danger? The environmental and health effects of nanoparticles are unknown.

of nanoparticles. The centre and associated research projects, which would cost £6 million (US\$11 million) a year to run, would look at the behaviour of nanoparticles in air, water and soil, as well as in cells.

The authors did not go as far as some environmental groups wanted. The ETC Group, based in Ottawa, Canada, for example, has called for a moratorium on nanotechnology research until more is known. But the group welcomed the report's precautionary tone, describing it as the "clearest recognition yet" of the need for new regulations.

War on polio back on track as Nigerian state restarts vaccinations

Washington The Nigerian state of Kano has resumed vaccinating children against polio after a break of almost a year, renewing hope that the crippling disease may be finally eradicated. During the past year, polio has spread from Nigeria to ten other African countries that had previously declared themselves polio-free.

Islamic clerics in Kano had boycotted the oral vaccine, claiming it carried HIV and was

laced with contraceptives to stifle the growth of the state's predominantly Muslim population (see *Nature* **428**, 109; 2004). Now Kano has found a new supplier of the vaccine in the Islamic nation of Indonesia.

The state's governor, Ibrahim Shekarau, launched the campaign by personally vaccinating several infants.

Gene-therapy pioneer charged with child abuse

San Diego The University of Southern California has been plunged into turmoil with the arrest, on charges of child molestation, of one of its most prominent scientists. On 30 July, gene therapist William French Anderson was charged with sexually abusing a young girl, whom he coached in taekwon do, over a four-year period from 1997 to 2001. As *Nature* went to press, Anderson had pleaded not guilty and was released on \$600,000 bail.

Anderson pioneered the field of gene therapy at the National Institutes of Health in Bethesda, Maryland, starting the first approved human clinical trial in 1990. He moved to the University of Southern California in 1992, and has remained active within the field. The university authorities have placed Anderson on administrative leave.

Condensed matter is a big hit with physicists

Washington Two papers on the arrangement of electrons in solids, both from the 1960s, top a list of the highest-impact physics papers published in the *Physical Review* journals over the past 110 years.

The papers in question were published by Walter Kohn and his colleagues in 1964 and 1965, and take the top spots thanks to the large numbers of citations they have generated — 3,227 in the case of the top paper. They are still cited frequently today. Kohn's work, carried out at the University of California, San Diego, has proved invaluable to researchers studying the theoretical properties of solids and liquids, and earned him the 1998 Nobel Prize in Chemistry.

Like other 'best of' lists, the rankings are likely to provoke debate. The list's author, physicist Sidney Redner of Boston University in Massachusetts, urges caution over its interpretation, as it covers only one publishing house. The journals are seen as the backbone of physics publishing, but many classic papers, such as Einstein's works on relativity, were published elsewhere.

Most of the top papers in Redner's list, including those by Kohn, are on the quantum properties of condensed matter.

♦ <http://xxx.arxiv.org/abs/physics/0407137>



Back on the menu: the oral polio vaccine will once more be given to children in Kano.



Dive in: amateur snorkelers are among those being encouraged to collect marine data.

Divers net information for sea conservation website

London Recreational scuba divers and snorkelers are to pool their observations of marine life in a new ocean conservation initiative. Backed by the UN Environment Programme's World Conservation Monitoring Centre in Cambridge, UK, Earthdive will ask divers around the world to look for key species, such as sharks, lobsters, marine turtles and butterfly fish.

Volunteers will also report the human impact on the sea by noting down the presence of litter and other pollution.

The results will appear on Earthdive's website, which the project's leaders hope will be a valuable resource for conservation biologists. "The greater value will come over time, as we can see long-term changes," says Earthdive director Angela Bawtree.

Hundreds of divers have already paid a fee of £10 (US\$18.30) to join the project, half of which goes to a selection of marine conservation organizations.

► www.earthdive.com

EPA pumps cash into heart disease study

Washington The US Environmental Protection Agency (EPA) has announced its largest-ever grant, to study the health effects of air pollution. The ten-year, \$30-million study, to be carried out at the University of Washington in Seattle, will look for correlations between particulate air pollution and cardiovascular disease in 8,700 volunteers over the age of 50 across the United States.

Exposure to airborne microscopic particles, from sources including vehicle exhausts, is associated with increased hospital admissions and deaths among older

people, largely because of lung and heart disease. The EPA expects the results of the new study to underpin its future regulation of particulate emissions.

Norwegian bids to broaden scope of science prizes

San Francisco The annual Nobel celebrations in Sweden may soon be joined by a comparable event in neighbouring Norway. Norwegian-born philanthropist Fred Kavli is planning a series of scientific prizes that will offer awards with a cash value similar to the Nobels, which are each currently worth about US\$1.3 million.

Three prizes are being planned, probably in the interdisciplinary areas of neuroscience, nanoscience and cosmology. "We want to be a little less conservative than the Nobel Foundation," says David Auston, president of the Kavli Foundation in Oxnard, California. Details of the planned prizes should be finalized by mid-September. They will probably be awarded every two years, beginning in 2007, and the foundation hopes to award them in Kavli's home country.

Kavli, a physicist, made his fortune through Kavlico, a company that makes sensors for aeronautical, automotive and other industrial applications.

► www.kavlifoundation.org

To the edge... and over



Special report on the Olympics

When the Olympic Games open in Athens on 13 August, all eyes will be on the athletes. But it's not just the spectators who will be showing an interest. Sports scientists and doctors will also be keeping a watchful eye on the competitors — to them, the Olympics are a kind of giant experiment.

Few elite athletes are willing to participate in lab-bound projects. So the games are an important component of trials that help to determine how well various training strategies work. Such regimes can include high-tech equipment to pinpoint exactly which muscles are used in a given event, strict diets, and high-altitude training (see 'A breed apart', opposite).

The Olympic Games showcase talented people pushing their bodies to the absolute limit. In the following pages, *Nature* looks at the unusual



problems faced by this elite group — beyond the normal wear and tear of sprained ankles, pulled muscles and torn tendons. For some, their bodies are worked so hard — simply from training, let alone the added abuse of any hormones or drugs — that they develop a syndrome of perpetual exhaustion (see 'The medals and the damage done', page 604). For others, a rare genetic condition can mean that their decision to compete in the games could prove fatal (see 'Heart-stopping action', page 606).

Researchers are also keen to see how the athletes will cope with Athens' heavy pollution. For cyclists and marathon runners, in particular, the events will be as much a test of their lungs as of their legs (see 'Gasping for victory', page 608).

Elite sport, a triumph of the human body over the laws of nature, pushes participants to the edge of possibilities — and sometimes right over that edge. In the coming weeks, *Nature* will be watching to see what records — and which athletes — are broken in the process. **Nicola Jones, assistant news and features editor**

For additional Olympics coverage ▶ www.nature.com/news/olympics

Gymnasts put extreme strain on their bodies for moves such as the Maltese cross (left).

need to look at genes and their responses to environmental triggers. "I'm convinced that if we could look at all the genes that are induced by low oxygen levels in a big enough pool of athletes, we would learn how to predict which ones would benefit from altitude training," says Hawley. For the moment, that kind of study seems out of the question.

Even if the genes that control responses to everything from oxygen levels to different nutrients could be pinned down, and every athlete provided with an ideal training schedule and diet, it wouldn't level the playing field.

Only those whose genes are preset for maximal athletic performance will ever join the elite band of Olympic athletes. Eero Mäntyranta, the 1964 Finnish cross-country skiing gold medalist, for example, had a mutation in the gene encoding for erythropoietin, a protein that regulates the production of red blood cells, which sports scientists

believe accounted for his extraordinary stamina. There are hundreds of other genes — from those determining body proportions to those optimizing oxygen and nutrient utilization by muscles — that help tip the balance towards élitism.

Although the elite athlete — packed with performance-enhancing genes and aided by the best technology available — is indeed a different kind of creature, in at least one way they are just like the rest of us: to improve, they need to work hard. As physiologists freely admit, the biggest impact on performance is still down to good-old-fashioned sweat. "What science can bring to the athlete is perhaps 0.1%," says Davis. "99.9% is still down to the athletes themselves."

The standards for hard work in sport have increased dramatically over the years. Roger Bannister — who in 1954 became the first person to run a mile in under 4 minutes — trained for only 35 minutes per day. That's barely considered a warm-up today, when the mile record stands at 3 minutes 43.13 seconds. To bring that down any further is going to take an awe-inspiring combination of extreme physiology and hard work. ■

Alison Abbott is Nature's senior European correspondent.

1. Levine, B. D. & Stray-Gundersen, J. *J. Appl. Physiol.* **83**, 102–112 (1997).
2. Ashenden, M. J., Gore, C. J., Dobson, G. P. & Hahn, A. G. *Eur. J. Appl. Physiol. Occup. Physiol.* **80**, 479–484 (1999).



Electromyography allows athletes to identify which muscles are active.

A breed apart

Olympic athletes are on the edge of normal physiology, finds Alison Abbott.

"The physiology, biomechanics and psychology of the elite athlete is something else," says John Hawley, sports scientist from the RMIT University in Melbourne, Australia. "Compared with the merely fit, that athlete is an entirely different animal."

Like all sports scientists, Hawley would like to understand why — but he knows he has chosen a difficult test species. Getting biological data out of elite athletes is almost impossible. By definition, they are in perpetual, intense training programmes that they do not wish to interrupt — either to take part in experiments or to donate blood or muscle samples. "It's not so easy to get an Australian cyclist who earns over a million dollars a year to stick a thermometer in his butt while training — you have to be able to sell your idea very convincingly," says Hawley.

And despite the number of dedicated institutes that have sprung up in many countries to support top athletes, sports scientists also know that they will have limited input into training strategies designed to hone individual performance. Coaches — usually former athletes themselves — tend to stick with training programmes that have been well tested in the field. More experimental ideas usually take a long time to be accepted by the sporting community.

The exceptions to this rule are advanced technologies — such as video feedback — that can help athletes focus their more traditional training. "Being able to track performance in real time has had a major impact on technique training in events such as high-jump and diving," says Peter Davis, director of coaching and sports sciences at the US Olympic Training Center in Colorado Springs. Electromyography (EMG), which measures electrical activity in working muscles, is another technique that has proven popular. One trainer, recalls Davis, brought a gymnast to the Colorado Springs centre for EMG analysis and was stunned to find that the muscle groups he had been focusing on for the 'Maltese cross' — a challenging move on the rings where a gymnast holds his body up in a horizontal position — were not the ones that the gymnast was actually using for the trick.

Such quick, definitive feedback is easy for trainers to work with. But more long-term, subtle studies are much harder. "We usually use recreational athletes to confirm that a particular training technique that has been used in the field actually works in the lab," says Hawley. But such studies are often inconclusive. Australian and US research teams have studied the effects of high-altitude training, for example — a technique long thought by trainers to improve performance. But the mechanism by which it works remains unclear. Some studies have shown an increase in the level of the oxygen-carrying protein haemoglobin found in red blood cells¹, whereas others have found no change in blood markers at all². And not everyone studied has actually benefited from the technique.

Sporting chance

To work out who benefits, and why, scientists

The medals and the damage done

Athletes' punishing schedules push their bodies to the limit. But what happens if they go too far? Jim Giles finds out.

The incessant drive for success can be a dangerous thing for an athlete pushing hard for the Olympics, as Michael Brennan (not his real name) discovered. In 2002, Brennan was a British national rowing champion. The following spring, he rarely finished outside the top three in any race. As the UK Olympic trials loomed, Brennan was feeling confident. But he is not going to this summer's games.

For much of the past 12 months, Brennan's performance has been eroded by constant colds, aching joints and fatigue. What seemed like minor infections refused to go away and extra training made matters worse. When the trials rolled round this April, Brennan still felt that he could qualify. Instead, he finished at the bottom of the heap. "I couldn't believe it," he says.

To an experienced sports doctor, the explanation is obvious: Brennan has 'unexplained underperformance syndrome' (UPS). Perhaps one in ten athletes will get the condition, but only now are hints emerging about the underlying causes. And although sufferers usually recover after a period of rest, it can ruin preparation for major events. "It's all over for an athlete if they get it before a competition," says Greg Whyte, sports science coordinator at the English Institute for Sport in Manchester.

Brennan was officially diagnosed with UPS after a visit to Richard Budgett, director of medical services at the British Olympic Association in London. Budgett, himself a rower and former Olympic gold medallist, sees about 50 athletes a year with the condition, and he says Brennan's story is typical.

Rowers seem particularly susceptible.



Working flat out: before competitions, rowers often force their bodies to adapt to intense exertion.

Athletes in this sport typically train twice a day, five or six days a week. When major competitions approach, they often push their bodies beyond their normal limits for one or two intense weeks, regularly covering 20 kilometres on the water a day. This 'over-reaching' leaves the athletes exhausted in the short term, but in the long term it usually forces their bodies to adapt to intense exertion, improving race performance. Usually. But sometimes the strategy results in UPS — also known as over-training syndrome. "There is a fine line between over-reaching and over-training," says Whyte. "When athletes bounce back from over-reaching they will be stronger, but some don't recover."

Feeling down last summer after a dip in form and problems in a relationship, Brennan responded by upping his training. "I felt I must do more," he says. Instead, continued infections and increasing fatigue forced him to do less. "I was feeling tired all the time," he recalls. "But I kept ignoring it."

These symptoms are typical of UPS, along with sleep and mood disturbances, loss of appetite, slow wound healing and gastrointestinal disturbances. If an athlete's performance does not recover after two weeks of relative rest, UPS is usually diagnosed¹.

Hidden menace

Gauging the prevalence of the condition is difficult, as few openly discuss it. "Athletes don't talk about injuries," says Budgett. "They don't want competitors to know." Brennan can be included in that camp, being happy to tell *Nature* his story as long as his real name wasn't used. Researching the condition is also difficult, as serious

athletes aren't willing to give regular blood samples, and it would be unethical to induce full-blown UPS in test subjects.

But, slowly, we are learning more about the syndrome. Over the past decade, a slew of studies has revealed numerous factors in the hormone and immune systems that seem to be linked with the condition, leading to tentative theories about the syndrome's cause.

One explanation was proposed last year by Lucille Lakier Smith, a sports scientist at the Tshwane University of Technology in Pretoria, South Africa². Her inspiration came from studies of victims of serious accidents,

such as burns patients. These patients have rather extreme immune reactions to deal with the severe damage to their tissues. At the same time, levels of some cytokines — molecules used to exchange signals between cells — increase. In burns patients, levels of anti-inflammatory cytokines,

such as the proteins interleukin-4 (IL-4) and interleukin-10 (IL-10), go up.

As well as calming inflammation, these molecules also suppress the ability of immune cells that usually attack pathogens. This leaves burns patients more vulnerable to long-lasting infections.

When athletes over-reach, Smith suggests, they similarly damage muscle, prompting an immune response. When she measured interleukin levels in athletes before and after a marathon, she found that IL-10 levels shot up, although IL-4 remained constant². This too could leave them vulnerable, she says.

But this is unlikely to be the whole story, as athletes with UPS often complain of tiredness and aching joints when they are apparently infection-free.

"After being ill I would train harder to make up for it. If I had backed off for a couple of weeks it could all have been sorted out."



Extreme events such as the marathon may provoke an immune response that makes the athletes tired.

A possible explanation for this comes from an academic who has personal knowledge of UPS. The condition hit Paula Ansley-Robson when she was attempting to break into the British Olympic rowing squad in 1996. "I was training for over 24 hours a week," she says. Then UPS struck.

"After that, a 5-kilometre run felt like a marathon."

Now an exercise physiologist at the University of Portsmouth, UK, Ansley-Robson is interested in another cytokine, interleukin-6 (IL-6), levels of which often rise when someone has a cold or flu. It makes the sufferer

feel tired, and is one of the signals that helps the body to slow down while fighting off an infection. It is also involved in regulating glucose energy stores during exercise. Levels of IL-6 can rise 100-fold during a marathon and are partly responsible for feelings of tiredness during a race.

Studies have shown that injections of IL-6 into healthy individuals will make them tired within a few hours. Some people are more sensitive to the cytokine than others. Chronic fatigue sufferers, for example, will become more tired more quickly, and for longer, than control subjects after injections of IL-6 (ref. 3). Injecting a runner before a race will likewise have tiring effects; an experienced club runner typically has a minute added to their time for a 10-kilometre run³.

Sensitive issue

Ansley-Robson suspects that over-reaching itself sensitizes athletes to IL-6, making them more likely to feel worse effects from natural bursts of the cytokine. She plans to put runners through a gruelling test of this idea. Runners will first do a timed 10-km run after an IL-6 injection. They will then 'over-reach' by running an incredible 90 kilometres — more than twice as far as a marathon. After a few weeks of rest, they will have an IL-6 injection and run a second 10-km race. If over-reaching sensitizes them to IL-6, they will perform particularly poorly in this last test, she predicts.

In the meantime, athletes can take some precautions. Michael Gleeson, a sports scientist at Loughborough University, UK, recommends a minimum six-hour rest between training sessions, as IL-6 levels are significantly higher among subjects who rest for only three hours⁴. He also emphasizes the need for coaches to monitor athletes' mood. Psychological stress is thought to contribute to UPS, although no one knows why.

For Brennan, careful monitoring of his mood and health could have made all the difference. "After being ill I would train harder to make up for it," he says. It's an understandable reaction from an athlete dedicated to the idea that training produces results. Yet Brennan now recognizes that less is sometimes more. "If I had backed off for a couple of weeks it could all have been sorted out," he says.

But he didn't. So when the rowers take to the water in Athens, Brennan will be watching on TV — although his thoughts have already turned to the 2008 Olympics in Beijing. For those games, he fully intends to qualify — after a good, long rest.

Jim Giles is a reporter for Nature based in London.

1. Budgett, R. *et al. Br. J. Sports Med.* **34**, 67–68 (2000).
2. Smith, L. L. *Sports Med.* **33**, 347–364 (2003).
3. Robson, P. J., Milander, L., Collins, M. & Noakes, T. D. *Can. J. App. Physiol.* (in the press).
4. Ronsén, O., Kjeldsen-Kragh, J., Haug, E., Bahr, R. & Klarlund Pedersen, B. *Am. J. Physiol. Cell Physiol.* **283**, C1612–1620 (2002).

Cooling off

When forcing their bodies through the trauma of extreme sport, athletes are grateful for any tricks that will help them to recover more quickly. For the Australians — an Olympic team that has turned the art of damage control into a real science — this means a strict regime of cold baths. "What we are doing differently this Olympics is a shift towards cold immersion," says Shona Halson, a fatigue and recovery scientist at the Australian Institute of Sport.

The Australian team has set up an entire health centre in Athens devoted to recovery, complete with mobile plunge pools. After an event, athletes will submerge themselves in water at about 12 °C for up to a minute, jump out for a minute, and repeat the cycle three to five times.

Most people who have done a hard day's sport know that an ice pack can calm swelling and

soothe aches. But does it do any real good for the body? Surprisingly, the evidence is largely anecdotal. Australian swimmers report that it cuts the time they have to spend doing warm-down laps by about 25% — they do these laps until their lactate levels are below a critical level, measured by pinprick blood samples by the side of the pool. But no proper studies on the effect have yet been done with the swimmers, simply because none of them will give up the baths to be in the control group — at least, not until the Olympics are over.

One theory about how cold baths work is that repeated constriction and dilation of blood vessels helps to flush out lactate — an acidic by-product of the body's combustion of sugars that can stop muscles from contracting properly. And, Halson says, the baths dull pain, giving athletes a psychological edge in subsequent events.

The icy baths used by the Australians may sound a little extreme, but they are not as bad as the 'cryochambers' popular with some European nations such as Germany and Poland. For these, athletes don socks, gloves and face masks before braving air at –110 °C to –150 °C for up to several minutes. Such chambers may or may not work better than ice baths, but they require constant medical supervision, says Halson.

The more casual sports-person may think that heat should be the way to unwind, but saunas and spas, says Halson, are definitely out. "They dehydrate the body," she says. And that's the last thing an overworked athlete needs. **Carina Dennis**



Cold comfort: ice baths may help athletes recover.

Heart-stopping action

Some athletes diagnosed with a genetic heart condition are banned from competitive sport. Laura Spinney discovers how they are fighting for the right to run the risk of death.

AFP

Domenico Fioravanti won two gold medals at the Sydney Olympics in 2000 — for the men's 100-metre and 200-metre breaststroke. But when he goes to Athens this year, it will be as a television commentator, not as a competitor.



His fans will have seen him most recently in the reality TV show *La Fattoria*, or *The Farm*. In fact, he hasn't been in a swimming pool at all this year — though not by choice. In January, the 27-year-old Italian swimmer was diagnosed with an inherited heart condition called hypertrophic cardiomyopathy (HCM) and, under Italian law, he was forced to retire from competitive sport.

Fioravanti suffers from the same genetic problem that killed Cameroonian footballer Marc-Vivien Foé in June 2003 while he was playing in the Confederations Cup semi-final against Colombia, followed seven months later by Hungarian Miklos Feher, who died while playing for Portuguese football team Benfica. Both men died on the pitch — suddenly, and in front of the cameras. Even though Fioravanti knows he has the same condition and is well aware of the risks, his family is campaigning for a relaxation of the Italian law. They want him to keep on swimming — or at least to have the option.

HCM is the leading cause of sudden death in people under the age of 30. It is caused by the accumulation of an abnormal protein inside sarcomeres, which are a basic component of heart muscle cells. This causes the cells to grow too large and in a shambolic fashion, particularly around the left ventricle. As the muscles thicken, the heart can develop an irregular beat and runs the risk of stopping completely. About 0.1% to 0.2% of the world's population has HCM; each year, about 1% of these die¹. The extra strain of excessive exercise is thought to trigger sudden death in people with the underlying condition².

Doctors can diagnose HCM through imaging studies, treat it with drugs or surgery, and implant a defibrillator to kick-start the heart back into action should there be a major problem. But it is still difficult to assess the risk of sudden death in any given patient. Fioravanti's frustration stems from this final

fact. To all appearances, his condition is mild. "At this point, I could state that he has a low probability of dying suddenly, even doing active sport," says Franco Cecchi, the cardiologist at the University of Florence who confirmed his diagnosis. Yet in Italy, the mere confirmation of the condition means that Fioravanti is banned from competitive sport.

Genetic testing is now gaining a foothold in studies of HCM, and holds the promise of sorting out this final detail. Fioravanti hopes that such tests may one day weed out those who have a high risk of suffering sudden death from those with a low risk, and so allow people like him to compete.

In Italy, uniquely, athletes are required by law to get an annual fitness certificate before being allowed to take part in any competitive event. As part of this assessment they are given an electrocardiogram (ECG) and questioned about their family history. If there is a suspicion of disease, they go on to have an echocardiogram ('echo') — an ultrasound scan of the heart and valves, which provides information about muscle thickness and the size of the heart's chambers. If this shows an abnormality, the athlete is automatically disqualified — unless the sport is considered low risk, such as snooker or archery. If an athlete sneaks through the system and goes on to die while playing, the doctor who signed the authorizing certificate can be held liable if found negligent.

Protect and survive

The Italian system seems to prevent instances of sudden death among athletes, says Cecchi. In 1998, Domenico Corrado, a cardiologist at the University of Padua, and his colleagues carried out a survey of people screened for the condition between 1979 and 1996. Of some 34,000 hopeful athletes, 22 had HCM and were disqualified from competitive sport. Only one athlete, who slipped through the tests, subsequently died of HCM³.



Miklos Feher (left) and Marc-Vivien Foé (pictured right) died on the field.

Cecchi admits that the system has its flaws. Corrado's team showed that screening is less efficient at detecting another potentially lethal heart abnormality called arrhythmogenic right ventricular cardiomyopathy. And it may be over-protective, banning people such as Fioravanti who could probably safely swim on. Nevertheless, the rate of sudden deaths among Italian athletes has been steadily decreasing, says Cecchi.

In the single case that escaped the screening net in Corrado's study, the clinical diagnosis was uncertain. Diagnosis of HCM is often not clear-cut, partly because exercise causes non-harmful changes to an athlete's heart muscle that can be mistaken for a sign of HCM. Some of this confusion might be avoided if a genetic test existed that could pinpoint the 200 or so mutations that have,

LUSA/AP (LEFT); R. DE LA MAUVINIERE/AP



Dicing with death: Domenico Fioravanti has two gold medals — and a genetic condition that has led him to be banned from competitive sport.

to date, been linked with the condition.

These mutations occur in 12 genes, all of which code for the sarcomeric proteins in heart muscle⁴. So far, genetic testing for HCM in Europe and the United States has been available only to individuals with a high risk of having HCM, and then only under the auspices of research programmes. But the first commercial test went on sale this May in the United States. Designed by Jonathan and Christine Seidman of the genetics department at Harvard Medical School in Boston, Massachusetts, the test is based on detecting mutations in eight genes that the couple think account for 75% of cases of HCM.

On the right track?

Barry Maron, director of the Minneapolis Heart Institute Foundation, has serious doubts about the usefulness of this genetic test for diagnosis. It misses not only 25% of the genes already linked to the condition, he says, but also the many other genetic mutations that are probably involved but have not yet been discovered. Even if a comprehensive test were available, he says, cost would prohibit its use for young athletes.

The Seidmans admit that the price tag of US\$3,000 means their test may not be feasible for screening the vast population of young athletes. Even ECG and echo tests are currently considered too expensive for many.

“But it might be worth considering for the considerably smaller population of Olympic athletes,” says Christine Seidman.

Yet it remains unclear how these genetic mutations affect risk. At the University of Florence, Cecchi’s team is studying some 350 patients and their families in an attempt to work out how, and to what degree, each of the HCM-related genes predisposes patients to sudden death. So far their work remains inconclusive⁵.

Even if the issue of diagnosis can be cracked, it still leaves policy-makers and athletes with the dilemma of what to do with the information. One case frequently discussed by doctors concerned about this issue is that of Nicholas Knapp. A talented 17-year-old basketball player, Knapp had already been offered a sports scholarship at Northwestern University, Illinois, when his heart stopped at the end of a game in 1994. He was resuscitated, diagnosed with HCM and surgically implanted with a defibrillator. Nevertheless, he accepted the sports scholarship. The university allowed him to keep the scholarship, but barred him from playing intercollegiate basketball on medical grounds.

A lengthy court battle ensued, in which Knapp argued that the choice was his, even if by playing he risked death. Northwestern’s decision was eventually upheld, but Knapp was free to enrol in another university’s basketball programme, and did so. Three years after his first cardiac arrest, he suffered an

episode of tachyarrhythmia — an abnormally fast and irregular heartbeat. His defibrillator discharged, probably preventing a second arrest and saving his life.

Since Knapp’s case, it has been hard for athletes with a known condition to play competitively in the United States. But, ten years on, there are still no formal laws to govern these decisions, only guidelines. HCM patients had to wait until this June for a guide⁶ on participating in recreational sport. The American Heart Association recommends that people avoid ‘burst’ activities such as sprinting, and rates the advisability of different sports for people with different conditions. Swimming is listed as “probably permitted” for HCM patients — although this is not meant to apply to athletes at the Olympic level.

Tough love

Maron, who helped to write these recommendations, says he thinks that the decision about whether a person with HCM can participate in sport should be taken out of the athlete’s hands. “You can’t just hand out freedoms against medical advice,” he says. His own research shows that the median age of HCM sufferers who die while engaged in sport is very young — just 17 years old.

In the United Kingdom, William McKenna, a cardiologist at University College London, describes the situation as “shambolic”. Athletes are not routinely checked for HCM. And when a sudden death occurs, he says, relatives are rarely informed about the victim’s genetic condition or the possible risk to themselves. Earlier this year the UK government agreed to establish new

“The median age of sufferers who die while engaged in sport is just 17 years old.”

guidelines on handling cases of sudden death. But the guidelines are likely to steer clear of any recommendations about screening or advice for young athletes.

While various countries battle with the ethical issues, the science marches on. If genetic tests become more conclusive, and heart scans cheaper, then at least we will have a better chance of diagnosing risk. But that is a distant hope for Fioravanti, who says he has resigned himself to the premature end of his career. He considered going to live in another country where he could continue to swim, but decided he would suffer too much away from home. So after the closing ceremony of this year’s Olympic Games, he will begin building his new career — as a trainer, rather than a swimmer. ■

Laura Spinney is a freelance writer based in London.

1. Maron, B. J. *Circulation* **106**, 2419–2421 (2002).
2. Corrado, D., Basso, C., Rizzoli, G., Schiavon, M. & Thiene, G. *J. Am. Coll. Cardiol.* **42**, 1964–1966 (2003).
3. Corrado, D., Basso, C., Schiavon, M. & Thiene, G. *N. Engl. J. Med.* **339**, 364–369 (1998).
4. Morita, H. *et al.* *Cold Spring Harb. Symp. Quant. Biol.* **67**, 383–388 (2002).
5. Torricelli, F. *et al.* *Am. J. Cardiol.* **92**, 1358–1362 (2003).
6. Maron, B. J. *et al.* *Circulation* **109**, 2807–2816 (2004).



Gasping for victory

This year's Olympics will take place in Athens, a city renowned for its hot, humid summers and stifling air pollution. Martin Leeb discovers how the athletes plan to overcome the hurdle of these tough conditions.

In 776 BC, when the first Olympic Games took place, Athens was a small town with a few thousand inhabitants. Things have changed. Nearly three thousand years later, the Greek capital is an industrialized, congested and chaotic sprawl of more than 3 million people, and one of Europe's most polluted cities.



The choice of Athens as the venue for this year's games must have come as a surprise to those concerned about environmental conditions at the site — including Olympic hopefuls. Dick Pound, former vice-president of the International Olympic Committee, revealed in his book *Inside the Olympics*, published this May, that concerns about the Athenian air quality led him to vote for Rome as the 2004 venue. "The evaluation commission failed to report on the air-pollution problems of the Greek capital," he wrote. "To this day when I hear the world 'Athens' my visceral memory dating back to my first visit is the stench of diesel fumes. The conditions in Athens made the much-vaunted smog of Los Angeles appear to be a breath of fresh air."

Concerns about athletes' health in such conditions are justified. During the 1984 Olympics in Los Angeles, British runner Steve Ovett collapsed after the 800-metres final with severe respiratory problems and had to spend two nights in hospital. "Pollution was one of

the major factors in my having exercise-induced asthma in Los Angeles," he told *Nature*. "There was a significant number of sufferers there but not much was reported."

In Athens, things have been improving. Ozone concentrations have decreased by about 13.5% in the past ten years, thanks to government initiatives, according to the local organizing committee for the games. Further measures are in place to reduce air pollution, including schemes to encourage spectators to use public transport rather than private cars.

Winds of change

"I don't think air pollution will be a problem," says Christos Zerefos, a climatologist at the University of Athens. But this is a bit of a gamble, as his prediction relies on seasonal winds that normally sweep the eastern Mediterranean clean from the end of May to the beginning of October. In August, this 'Etesian wind' is usually active for 6–20 days — but if it blows lightly this year, some say that the city could be in trouble.

Two pollutants of particular concern to athletes are ozone and particulates — tiny bits of soot and other matter — both of which can irritate the lungs and cause asthmatic symptoms^{1,2}. In Athens, many August days are char-

acterized by a brownish–yellow smog of photochemical pollutants, which the locals call 'nefos'. When the nefos hangs, ozone levels can rise to more than $160 \mu\text{g m}^{-3}$, some $40 \mu\text{g m}^{-3}$ higher than the World Health Organization's recommended limit. Randy Wilber, the US Olympic Committee's senior sports physiologist, measured ozone highs of $240 \mu\text{g m}^{-3}$ around the Acropolis last summer. At the same time, particulate concentrations can hit $50\text{--}60 \mu\text{g m}^{-3}$ — just above the $40\text{--}50 \mu\text{g m}^{-3}$ limit set by the European Commission.

Exactly how particulates irritate the lungs is unclear, but it seems that coatings of tiny amounts of metals such as iron or copper on flecks of dust and soot can trigger reactions in the lungs that generate highly reactive free radicals³. That can lead to inflammation of lung tissue, tightening of the chest and shortage of breath. In severe cases, fluid may enter the lung, requiring emergency treatment.

Athens isn't the only place with a pollution problem. Cities such as Los Angeles and Beijing have similar or worse conditions, causing a small background increase in breathing problems in the average population. During the 1984 games, ozone levels hit highs of $500 \mu\text{g m}^{-3}$ in the Los Angeles basin, contributing to Ovett's collapse.

Some 7% of the general population suffer from asthma. Heavy exercise can exacerbate the problem and about 13% of normal people and 20% of athletes report asthma-like symptoms as a result. This figure shoots up to 50% for road cyclists, who tend to breathe heavily during their endurance event. The

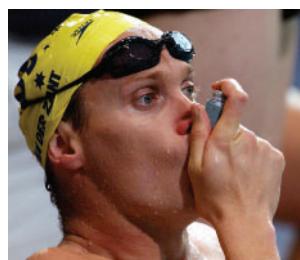
Predictions for Athens in August

Ozone-level high: $160 \mu\text{g m}^{-3}$

Particulate-level high: $50\text{--}60 \mu\text{g m}^{-3}$

Average daily temperature high: 33°C

Humidity: 49%



Deep breaths: clockwise from top left, Steve Ovett, whose performance in Los Angeles was hampered by pollution; and Paula Radcliffe, Jan Ullrich and Robert van Der Zant, all of whom suffer from asthma.

AP

list of afflicted athletes includes Paula Radcliffe, who holds the world record for the women's marathon, and cyclist Jan Ullrich, who won the Tour de France in 1997 and a gold medal at the Sydney Olympics in 2000.

The extent to which athletes are likely to be affected by pollution depends to some degree on their event. Endurance and outdoor athletes are more at risk than, say, table-tennis players. In Athens, road cyclists are likely to be most affected, says Wilber, because the races will take them close to the Parthenon and the Acropolis — one of the most polluted districts of the city.

Hard to endure

Of the two airborne irritants, ozone is the most problematic, says Frank Kelly, a professor of environmental health at King's College in London. For endurance events such as road cycling or marathon running, the threshold of ozone levels that could cause lung tissue to inflame will easily be exceeded if the cleansing Etesian wind isn't blowing, says Kelly.

Athletes can cycle about 150 litres of air in and out of their lungs every minute during competition — ten times more than normal — which exposes them to more pollution. To make matters worse, they also inhale much more deeply, taking pollutants into the deepest regions of the lungs. "Even spectators who are exposed to these pollutants are easily taken up to the threshold level," says Kelly. "But it is the athletes who are at greatest health risk."

Even if few athletes are likely to need hospital treatment, the air pollution in Athens will probably affect their performance. Few studies have been done to assess the extent of pollution's effects — and most of the work has involved recreational athletes rather than the elite. But the data that do exist are worrying. Greg Whyte, sports-science coordinator at the English Institute of Sport in Manchester, UK, has exposed highly trained athletes to ozone concentrations of $100 \mu\text{g m}^{-3}$ — over half the

level expected for a hot and calm August day in Athens. His unpublished results show that performance decreased by about 3–4% — a lot for athletes who can win or lose events by a fraction of a second.

Tough climate

National Olympic teams have been trying to make the best of the situation. The US Olympic Committee took some of its athletes to Athens to see how they would perform. "We measured their respiratory functions before and after test competitions at some of the Olympic venues," says Wilber. About 20% of the team members are known asthmatics, or are known to suffer from exercise-induced asthma. But a further 30% of the non-asthmatic athletes were sensitive to Athens' air-pollution levels. The difference was subtle, but for these athletes, the pollution made breathing just a little harder.

Other national committees do not consider such tests particularly wise. "Training in an industrial area and deliberately exposing athletes to a risk is not a rational option," says Hans Holdhaus, director of the Institute for Sports Medicine and Science in Vienna, who will look after Austria's athletes in Athens. Instead, the Austrian trainers have

limited themselves to telling their athletes about the dangers of pollution, and have advised them to restrict their training when ozone levels are high.

But knowing which athletes are likely to have a problem with the conditions is crucially important, says Wilber, as they can then be treated with anti-asthmatic drugs. The most effective of these, including a class of drugs known as β -agonists that help to relax the airways, are on the Olympic banned list and are available only to proven asthmatics. But other sensitive athletes can take alternative anti-inflammatory medication, including leukotriene antagonists — a new class of anti-asthmatic drug that blocks receptors in the lungs that normally trigger inflammation — or antioxidizing agents such as vitamins C and E. They might not be as effective as β -agonists, but at least they are legal. And

Whyte's studies show that they help to protect against the inflammatory effects of air pollution by mopping up noxious free radicals before they can do any harm. "Now we are able to adjust and optimize athletes' medication," says Wilber.

Heat will be another issue. Experts are predicting that Athens could win gold for the hottest games ever held — a dubious honour currently held by St Louis, the 1904 Olympic host, where the average daily maximum temperature in summer was 31°C . At those games, only 14 of the 34 starters finished the marathon. In Athens this August, the average daily maximum temperature is expected to be more than 33°C , with an average humidity of 49%. "We anticipate that there will be several collapses due to heat," says Wilfried Kindermann, head physician for the German Olympic team. In the worst cases, true heatstroke can be a life-threatening emergency, he says.

Hot on the heels of pollution tests, the United States has also sent its athletes to islands such as Mallorca and Crete to try to acclimatize them to the Mediterranean climate, says Wilber. Acclimatization to such conditions seems to prevent heat injury⁴. But even with the best preparation, athletes will lose fluid more quickly and will be more prone to collapse.

Coaches will be watching to see who will be affected most by the stresses of heat and pollution. And not just this year. The 2008 summer games in Beijing will likewise be a challenge for athletes. The annual average particulate concentration there is as high as $162 \mu\text{g m}^{-3}$ — about three times that in Athens — and temperatures are similar to those in the Greek capital. Athletes: start acclimatizing now. ■

Martin Leeb is an intern with Nature in Munich.

1. Mudway, I. S. & Kelly, F. J. *Am. J. Respir. Crit. Care Med.* **169**, 1089–1095 (2004).
2. Brunekreef, B. & Holgate, S. T. *Lancet* **360**, 1233–1242 (2002).
3. Donaldson, K. et al. *Free Radic. Biol. Med.* **34**, 1369–1382 (2003).
4. Martin, D. E. *Sports Med. Train. Rehabil.* **6**, 249–267 (1996).

V. R. VIEIRA/PA

N. SIMPSON/EMPICS (LEFT); L. REBOURS/AP

Traumatic events take their toll on mental health

Results of violence and natural disasters could overwhelm services in developing world.

Sir—Your recent News Feature “Asia’s tigers get the blues” (*Nature* **429**, 696–698; 2004) notes that levels of depression and suicide are rapidly increasing in East Asia. The breakdown of traditional community structures, the loss of family-support systems, long-distance immigration and economic uncertainty all take their toll, in the form of a rise in mental disorders. As noted in the News Feature, this cocktail of psychological stresses leads to depression, substance abuse, eating disorders and other stress-related problems.

There is, however, another important reason for the rise of depression and other mental disorders in China. During the past year, almost half a billion people there have

been affected by natural disasters such as floods and earthquakes, which have caused more than 2,000 deaths and the displacement of some seven million people, according to the World Health Organization (see www.who.int/disasters/country.cfm?countryID=10&DocType=2).

As the number of disasters — both natural and manmade — increases around the world, we can expect to see more stress-related disorders in the twenty-first century. Post-traumatic stress disorder (PTSD), in particular, can be caused by natural disasters or by the wars and acts of violence that afflict many parts of the world. Traumatic events do not discriminate on the grounds of culture,

race, gender or age (see, for example, I. Derluyn *et al.* *Lancet* **363**, 861–863; 2004). And people exposed to such events are far more likely to develop psychiatric disorders such as substance abuse, major depression, PTSD and psychosomatic illness (P. P. Schnurr *et al.* *Science* **303**, 168–169; 2004).

The impact of PTSD is especially devastating in developing countries, where health services are helpless to cope with the flood of traumatized people seeking refuge and help.

Menachem Ben-Ezra

Department of Psychology, Sharet Building,
Tel Aviv University, Ramat-Aviv,
PO Box 39040, Tel Aviv 69978, Israel

OvaCheck: let's not dismiss the concept

Sir—Your News Feature “Running before we can walk?” (*Nature* **429**, 496–497; 2004) highlights concerns about a new proteomic test for ovarian cancer known as OvaCheck. This test measures breakdown products of proteins in blood serum and looks for patterns that may indicate disease. The appeal of the method, now under scrutiny, derives from its simplicity and potential power; the total number of breakdown products could eventually exceed the number of gene transcripts and proteins, and may therefore provide a subtler correlation between biological events and disorders than regular expression genomics and proteomics. This approach should not be dismissed before it has been fully investigated with the best possible technological means and expertise, and under rigorously standardized conditions.

Experts quoted in the story cite concerns about mass spectrometry, biostatistics, and also environmental, dietary and psychological factors that could produce artefacts in the data. In our experience, serum preparation, handling and storage could be the largest source of artefact in mass-spectrometry readings, especially for archived samples. We are currently investigating the effects of, for instance, different blood-collection tubes and the number of freeze/thaw cycles.

The whole concept is now in danger of being declared null and void, both by some who previously embraced it and by others who lack the adequate experimental experience to participate in the debate. Although unbiased, your News Feature is

likely to embolden the critics and reinforce the bleak view of this particular type of proteomic analysis, and perhaps of proteomic profiling in general.

No final assessment can be made at this point of the concept underlying the OvaCheck test. We hope that leaders from the clinical and analytical-chemistry communities will come together and make a concerted effort to refine the procedures before testing them in carefully designed clinical studies. Only then can we establish whether this approach can disclose hidden patterns of disease or whether it is the molecular equivalent of the Emperor's new clothes.

Josep Villanueva, Paul Tempst

Protein Center and Molecular Biology Program,
Memorial Sloan-Kettering Cancer Center,
1275 York Avenue, New York, NY 10021, USA

OvaCheck: doubts voiced soon after publication

Sir—Your News Feature “Running before we can walk?” (*Nature* **429**, 496–497; 2004) summarizes recent developments in the use of serum proteomic patterns, in particular the OvaCheck test, for early cancer detection. It states that “the first criticisms of OvaCheck hit the public domain in June 2003”.

Unfortunately, this is not accurate. Following publication of the first paper by Petricoin and colleagues on ovarian cancer (your ref. 1), *The Lancet* published several correspondences critiquing the method in July 2002 (*Lancet* **360**, 169–170; 2002). More importantly, the limitations of this approach are not restricted to those revealed by the new bioinformatic analyses

described in the News Feature. (For more information, see E. P. Diamandis *Mol. Cell Proteomics* **3**, 367–378; 2004.) The cautionary notes of many authors on the limitations of this technology are useful in tempering the original optimism that this method will revolutionize the way we diagnose cancer in the future.

Eleftherios P. Diamandis

Department of Pathology and Laboratory Medicine,
Mount Sinai Hospital, 600 University Avenue,
Toronto, Ontario M5G1X5, Canada

Editing for posterity gets red card in Real world

Sir—I read with amusement your News story “‘Inspirational’ leader quits Madrid heart project” (*Nature* **430**, 127; 2004). Reporting a comment that the Spanish government should pay more to attract high-profile scientists, as Real Madrid did to attract David Beckham, you explained “Beckham is a footballer who left Manchester United for Real Madrid in 2003”.

After discounting the possibility that newly arrived aliens were the target audience, I conclude that either the article's editor needs to get out more, or there is an institutional belief that serious subscribers to *Nature* cannot possibly find time for a life between the weekly arrivals of your journal, or both.

Derek Partridge

Department of Computer Science, Harrison
Building, University of Exeter, Exeter EX4 4QF, UK

correspondence

Contributions to Correspondence may be submitted to corres@nature.com. They should be no longer than 500 words, and ideally shorter.

Sex under pressure

If sex is so important for evolution, why don't we have more of it?

Why We Do It: Rethinking Sex and the Selfish Gene

by Niles Eldredge

W. W. Norton: 2004. 224 pp. \$24.95

Robert Foley

I opened this book with considerable anxiety. The blurb and advance praise on the back make clear that *Why We Do It* (have sex, that is) aims to shatter myths, recast darwinism and fundamentally change the way we understand our own evolution. Niles Eldredge, an invertebrate palaeontologist and one of the major figures in the macroevolutionary debates of the 1970s and 1980s, sets out to tear down the whole edifice of reproduction-driven neodarwinian behavioural and evolutionary ecology. If Eldredge's explicit idea is correct, then anyone reading this book should emerge completely purged of the orthodox model of evolution as a process of enhancing reproductive success. No wonder I was anxious.

Eldredge's argument is straightforward. Sex for most animals, including or even especially humans, does not happen very often. Most of life is filled with growing up, finding enough to eat and avoiding predators. Sex is only occasionally interspersed among these activities — I leave it to other readers to quantify this, as I suspect there must be considerable variation. Indeed, this is one of Eldredge's main points: there are some individuals — and we're talking about humans here — who never have sex. What follows from this, Eldredge argues, is that because sex is relatively rare, it must, in evolutionary terms, be relatively unimportant.

The next step in the argument is that sex is the source of the 'selfish gene' or 'gene-centred' model of evolution, so this model clearly must be wrong. Selection is not for sexual and reproductive success alone, but affects all the other events between birth and death. As Eldredge says, life is for living, not for having sex and reproducing, so fitness for life — not for the ability to spread genes — is what the game is all about.

Eldredge then sets out to establish that there are, within the arena of evolution, two separate spheres: reproduction and economics (drawing on some of Darwin's original formulations). Selection operates on both of these independently and, if anything, selection for economic factors is more important than selection for reproductive success, on the grounds that more effort goes into it. From this point comes a battery of assaults on the role of evolution in human behaviour: the dismissal of evolutionary psychology, the dominance of culture over biology,



Life is for living: selection affects every aspect of our lives, so does it all come down to sex?

the separation of sex from reproduction in humans (reducing even more the significance of reproductive success), and an attack on biological determinism in human affairs.

At heart, the book is deeply anti-gene. It argues that genes are wrongly placed at the centre of evolutionary theory, that genes do not play an active role in the evolutionary process, that genes do not greatly influence human behaviour, and that the genetic heritage of human evolution is not significant.

Is this the end for those Eldredge portrays as the ultradarwinists — notably Richard Dawkins, E. O. Wilson, all evolutionary psychologists and probably most behavioural ecologists? Has Eldredge demolished gene-centred evolutionary theory? Hardly. The logical error in Eldredge's argument is so breathtakingly obvious that one can only wonder how he has missed it. Reproductive success is not just about having sex. Take virtually any animal, and certainly it may spend relatively little time actually copulating, but what Eldredge isolates as the economic sphere is clearly not independent of sex and reproduction. How an organism grows — with males of many species being larger than females, for example — is not an isolated biological phenomenon, but represents an individual (through its genes) positioning itself to compete well in the reproductive arena. Its social behaviour, being competitive or cooperative, and its feeding ecology are all part of its reproductive strategy — and hence are selected not just for their own efficiency, but for the extent to which they contribute to reproductive success. Countless studies across vast

numbers of species have shown how patterns of 'economic' behaviour are related to reproductive strategies and success.

Humans are the central concern for Eldredge. It is to his credit that his primary argument is not that selfish-gene models do not apply to humans because they are different, but that selfish-gene models do not work for any organisms, and therefore apply even less to humans. Ultimately, as the final chapters make clear, this book sees in classic neodarwinism the dangers of biological determinism, and views evolutionary psychology in particular as the modern version of the older threats of social darwinism and eugenics. There is plenty on humans here, but whether this has much to contribute to debates on the relative roles of genes and the environment in human behaviour, and on whether human behaviour is shaped by strategic concerns over reproductive fitness, is a matter of doubt. The evidence on human behaviour takes the form of anecdotes about how we do not have as much sex as we could, or how we sometimes behave in ways that do not obviously seem to enhance our fitness. Although such anecdotes are indeed abundant, they hardly constitute serious scientific study.

It may seem harsh to criticize so heavily a book written in such an extraordinarily popular and friendly manner — this is not a scientific monograph, after all. However, the claims made here are so strong, so polemical and so tilted towards making reproductive fitness seem like an irrelevance in the evolutionary process that it would be inappropriate not to point out the extent to which

a naive reader might be misled. There are many oversimplifications and difficulties with the strongly adaptive models of human evolution constructed by evolutionary psychologists and behavioural ecologists, but Eldredge's approach is too extreme to bring these out in any way that might usefully influence future developments.

Should you read this book? It is written in a chatty and homely style, which will appeal to some and grate on others. Even the notes provide little bibliographic support for the points made in the body of the book, being rather an extension of the polemic by other means. If you are in the mood for some relentless Dawkins-bashing, or want a rush of arguments against biological determinism, then you might enjoy it. But if you are a neodarwinian in search of a road-to-Damascus experience, this is not the place to find it. And if you are a neodarwinian not looking for such an experience, you had better avoid this book, as its superficiality, inconsistency and misleading logic will only irritate. On the other hand, you need not have any fear of having your evolutionary world turned upside-down. I closed the book with a sigh of relief.

Robert Foley is in the Leverhulme Centre for Human Evolutionary Studies, University of Cambridge, Downing Street, Cambridge CB2 3DZ, UK.

The birth of modern science?

The Forgotten Revolution: How Science Was Born in 300 BC and Why It Had To Be Reborn

by Lucio Russo (transl. Silvio Levy)
Springer: 2004. 487 pp. \$99, £69, €89.95

Mott Greene

The Forgotten Revolution is a work of passionate advocacy, defending the excellence of science in Greece in the 200 years after the beginning of the Hellenistic period in 320 BC. The author, Lucio Russo, a professor of mathematics at the University of Rome, has an appetite for classical works of science in their original Greek and Latin, and is on a spirited crusade against two groups. His first target are those "moderns" who (apparently) lump everyone from the Greek philosopher Thales in the sixth century BC to Ptolemy, an Egyptian astronomer of the second century AD, into an undifferentiated mass called the Ancients, and then disparage their intellects and scientific achievements. The second group are the classicists, who are accused of treating everything that came after Athens in the fourth-century BC, and the Hellenistic world in particular, as a technically proficient but intellectually



Just a regular floor? Hellenistic design often drew on geometry.

uninteresting successor to the classical world.

If this is a fair portrayal of these groups then Russo is right to take up the cudgels, for the Hellenistic period is a brilliant episode in the history of science, and our own mathematical and physical sciences are built on its foundations. The connections between Russo's "forgotten revolution" and the scientific revolution of the seventeenth century are immediate and direct. The surviving treatises of Hellenistic mathematics, translated from Byzantine codices in the late sixteenth century, were essential to Descartes, Galileo, Kepler and Newton.

Russo presents an enticing vision of a Hellenistic world with a highly organized scientific effort that lasted for centuries. He sees government support for science, libraries and institutes. He sees science with technological applications, if not an industrial scale. He views the Hellenistic development of geometrical proof into a hypothetico-deductive method as the revolutionary birth of true science. If there is anything you like in modern science and mathematics (or art, linguistics, architecture, technology or medicine), he is glad to show you that the Hellenistic Greeks had already been there and done that. He is even pretty sure they had the inverse-square law of gravitation.

Is this vision true to the facts? A great deal of it is, yes, but not all, and certainly not the bit about the inverse-square law. Yet the effort to spin out the "what if...?" is well worth it. We have access to only 1–2% of these ancient texts for which we know the titles, the rest being lost, which leaves a good deal of room for Russo to imagine a Hellenistic science much more ample and more modern than previously thought. There is a corresponding tendency to downplay the quality and quantity of science before the fourth century BC,

and to find much after the second century BC decadent. The argument is nonetheless stimulating.

Fortunately for the reader (although more or less in contravention of the book's central premise), Russo no sooner limits science to hypothetico-deductive reasoning than he drops that stricture. This lets in lots of wonderful science, technology and medicine, achieved in the Hellenistic period by observation, description, approximation and induction. If we were stuck with his stricter definition of science, then almost everything in the Hellenistic world, and certainly everything in this issue of *Nature*, would not count as real science.

Russo is entitled to his own opinions but not to his own facts. There are some errors and bold exaggerations: *caveat lector*. His attempt to suggest that Ptolemy's

Almagest "presents a system for predicting the motions of the planets but no explanation of how the system was obtained" is both untrue and unfair, although Russo's thesis that Ptolemy belonged to a decadent age requires it. Ptolemy actually provided a beautiful, original, rationally organized description of all the key parameters, including how to derive them from observation, and how to obtain these observations under minutely specified conditions. The interested reader is referred to James Evans' *The History and Practice of Ancient Astronomy* (Oxford University Press, 1998).

Indeed, much of the history of science referred to by Russo resides in standard works published over the past 50 or so years. It seems he has not recently renewed his acquaintance with the subject through the technical literature, although he has read widely among the popular works. This does not undo the worth and verve of his presentation, but the sheer volume of recent work does undo his 'conspiracy of silence' thesis. Try using Google to track down a Greek name — Aristarchus of Samos, say — and you'll see what I mean.

Russo's book left me feeling that I would have loved to live (and might well have found a good job) in third-century Alexandria. The copious, informative and useful illustrations in this beautifully produced volume intensify the sense of what was lost when this great civilization went into decline and perished. They also inspire a sense of where we might be, had that "lost revolution" continued unchecked from 2,000 years ago to today. Would we be travelling to the stars? "One may imagine..."

Mott Greene is in the Department of History, University of Puget Sound, Tacoma, Washington 98416, USA.

..... Into the unknown

The Invisible Century: Einstein, Freud, and the Search for Hidden Universes

by Richard Panek

Viking: 2004. 288 pp. \$24.99

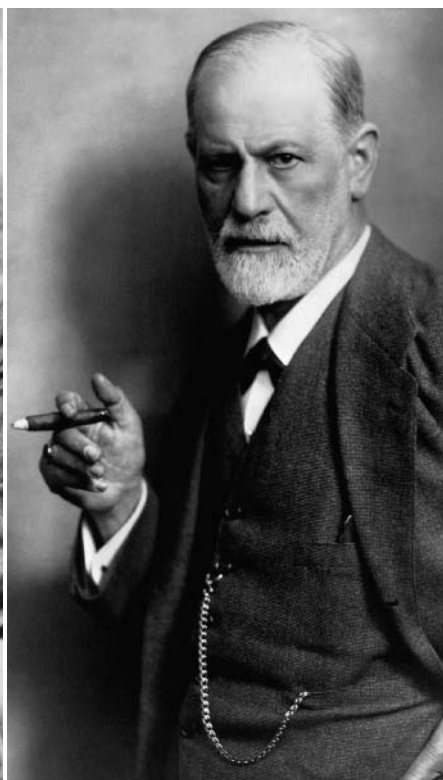
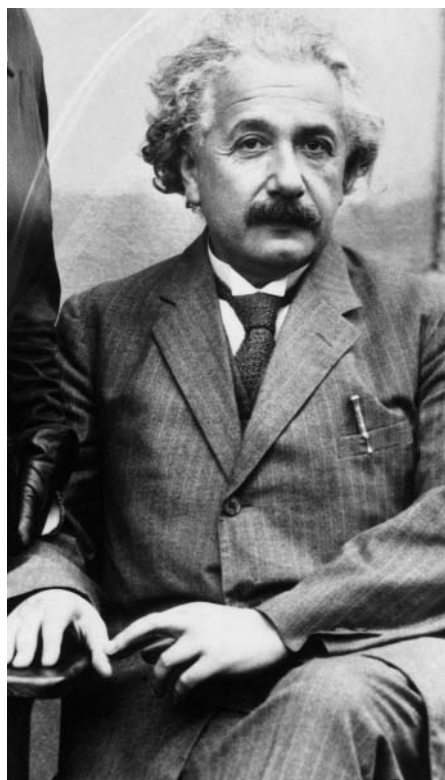
Diana Kormos Buchwald

The title of this book is a little ambiguous: science writer Richard Panek surely meant it to refer to the 'century of the invisible'. Following his earlier account of Galileo's discovery of previously unseen objects in the sky, *Seeing and Believing* (Penguin, 1998), he now turns to two giants of twentieth-century scientific and intellectual history, Sigmund Freud and Albert Einstein. He considers how they redefined the way we think about hidden entities and phenomena, in both the deepest recesses of our mind and the farthest reaches of a warped, infinite and expanding Universe.

The telescope, the microscope and other technological advances of the scientific revolution have allowed us to explore both macroscopic and microscopic objects that were previously invisible. Here Panek tells us how Freud and Einstein enabled modern man to rethink the way we think—a claim to an even more profound scientific revolution. Freud's psychoanalysis not only established a method of medical treatment for mild psychological ailments, but sought to understand and explain both the neurological and the psychological structure of the mind. Einstein, by postulating the constancy of the speed of light and the equivalence of the laws of physics in all systems in motion, forced us to rethink, recalculate and eventually abandon some long-entrenched fundamental assumptions about the structure of the Universe, such as the existence of absolute space and time. He led us to understand gravity and mass not merely as newtonian categories but as features inherent in the fabric of his new concept of space-time.

Panek's brief double biography, part history and philosophy of science, part essay on progress and civilization, comes at a time when Einstein is capturing the publishing, popular-science and academic limelight. Exhibitions in the United States, Europe, Israel and elsewhere, as well as conferences, books and articles, have already been devoted to the centennial anniversary next year of Einstein's ground-breaking early papers on special relativity, brownian motion and the photoelectric effect. Some 20 books on Einstein are due to be published in the next year.

Given the complexity of Einstein's scientific work, the general trend of recent Einstein biographies has been to explore his 'human side'. To accomplish this, some authors have resorted to cutting the man down to size—to the scale of folk percep-



Hidden talent? Albert Einstein (left) and Sigmund Freud both managed to explain the unseen.

tions of morality, goodness, frailty and failure. We have recently learned much about Einstein's illegitimate daughter, his turbulent divorce from his first wife, his extramarital affairs and his inclination not to quote predecessors in scientific papers. A similar treatment was apportioned to Freud in the 1970s. There were many scholarly, as well as journalistic, exposés of Freud's vanity, his competitiveness, his ruthless handling of colleagues, and, most famously, his views on the centrality of fantasy and invention in the sexual-abuse reports of his women patients.

Fortunately, Panek is, if anything, trying to bring some perspective and factual background to the principal achievements of Freud and Einstein, sparing us the moralizing rod. He insists on, and demonstrates, the complexity of reconciling theoretical, mathematical and psychological hypotheses and theories with empirical observations, a profound tenet of both scientists' methodology and creative work.

Both were trained at a time of exhilarating new discoveries in physics, astronomy, medicine and neurology. Both concentrated their efforts on reconceptualizing topics and structures that seemed well explored at the time. And both succeeded in making substantial progress, not by merely seeking answers to unsolved problems, but by asking new questions of old or overlooked material. Why do we routinely misplace keys and umbrellas? What would it be like to travel on a ray of light? Why do people laugh at jokes? How do we know whether we are moving upwards or standing still in an elevator?

What are dreams? What is light? These seemingly trivial fantasies and queries not only allowed the two scientists to escape from the "merely personal", but provided a lever for rearranging long-established experimental and clinical data. Panek covers a great stretch in the history of physics and sciences of the mind, including the concepts of ether and light, X-rays and the mathematization of modern physics. Also there are the discovery of the neuron, the role of hypnosis in Freud's emerging analytic techniques, and his struggle to formulate the analysis of dreams and their relation to the unconscious.

As Panek points out, even though Freud and Einstein corresponded, most notably on the causes of war, and deeply respected each other, they never felt that they were kindred spirits. Freud, in a by now well-known anecdote, reported that, because neither knew much about the other's subject, his meeting with Einstein in 1927 was pleasant enough and free of grandstanding. Panek's book will surely bridge this gap in understanding. It will provide physicists with a cogent and fluid introduction to the early history of psychoanalysis, and to an empathetic reading of Freud and his contemporaries; and its limpid prose will help psychoanalysts, psychologists and modern neuroscientists to understand some of the more complex developments in twentieth-century physics. Both men would have been pleased. ■

Diana Kormos Buchwald is associate professor of history, and general editor and director of the Einstein Papers Project, California Institute of Technology, Pasadena, California 91125, USA.

Science in culture



Shaping up at Olympia

Geometry lies at the heart of sculptures from the Temple of Zeus.

Stefano Grillo

The Temple of Zeus at Olympia was erected in around 450 BC, just a few paces from the stadium that was home to the original Olympic Games. The temple's eastern pediment was adorned with sculptures that tell of one of the main events of the competition: the chariot race.

According to legend, Oenomaus, king of Olympia, was told by an oracle that he would perish at the hands of whoever married his daughter, Hippodameia. So he decided to challenge each of her numerous suitors to a chariot race: only if they won could they marry his daughter. This was a shrewd plan, because his horses were a gift from the gods and hence were invincible. But Pelops, one of the suitors, was even shrewder: he bribed the king's charioteer, Myrtilos, to replace the screws holding the wheels with ones made of wax. During the race, the wax melted and the king was defeated and killed. Pelops then married Hippodameia, thus fulfilling the oracle's prediction. The region where Olympia is located (the Peloponnese) was named after him.

The pediment (shown at the top) is now preserved in the Archaeological Museum at Olympia, which re-opened in spring after renovation. The pediment's simple, sober composition shows the characters before the race. Zeus, appropriately placed at the centre, separates the self-assured Oenomaus and his wife on the right from the bold Pelops and his future wife Hippodameia, whose rigid pose is brought to life by the angle of her head. Nobody knows how the race will turn out except for an old seer — in whom the dignity and venerability of old age are magnificently rendered — sitting behind the horses on the right.



A sense of balance: Piero della Francesca used symmetry in the *Madonna del Parto*, in an echo of Greek sculptures from the Temple of Zeus (top).

These remarkable sculptures were created at a momentous time in the history of art: the early Classical period. The cultural achievements of this age have been eloquently termed "the Greek miracle". The stones seem to have a calm, serene life such that, far from being paler copies of reality, they communicate to us a heightened, fuller sense of it. The majestic, frozen poses of the figures — who, apart from the seer, betray no emotion — are reminiscent of works by the Renaissance painter and mathematician Piero della Francesca. In both artistic styles, geometry plays an important role.

The geometry of the pediment's composition is apparent both in its overall symmetry and in the details. The gradual rotation of the heads of the

four horses at regularly increased angles, more clearly visible in those on the right than in their damaged mirror image on the left, serves to enhance the sense of order, simplicity and enigmatic stillness of the scene. This rotation is balanced by the kneeling poses of the two figures in front of the horses, placed in mirror symmetry; their legs are raised at angles that would have pleasingly echoed the lines of the top of the pediment. There is another figure in a similar pose on the left. A young man sitting at the far right (not shown) — one of the best sculptures in the group — is a variation on this theme, with his position turned by 90 degrees towards the observer.

Piero used similar geometric operations in his compositions. The positions of his characters are often obtained by translations and rotations at different angles of the same human figure — and the figures in turn are generated from geometrical bodies. As for mirror images, on one occasion Piero used both sides of the same preparatory drawing to draw the contours of two symmetrically placed angels.

Of course, geometry by itself does not make good art — but it can powerfully enhance a work's meaning if it springs from the same source as the work. In Piero's world, geometry expressed the divine order of God's creations. In the Olympia pediment, geometry contributes to the harmony and serenity of the composition. These are the harmony and serenity of the Greek Classical world; they have inspired many subsequent artists but have never been equalled.

Stefano Grillo is a physicist at the University of Perpignan and PROMES/CNRS, Rambla de la Thermodynamique, Tecnosud, 66100 Perpignan, France.

The pleasure of learning

How an Italian visitor rekindled the joy of science in a war veteran.

Leon Lederman

Picking up the threads of normal life after a war is difficult, even when you are on the winning side. In 1946, returning from the Second World War, I registered at Columbia University, New York, as a graduate student in physics. That ensuing year was the worst that I can recall. The classes were crowded with veterans returning from military service or from research in the various laboratories that had supported the war.

The teaching faculty was also just beginning to return, and so most of the courses were taught by Isidor Isaac Rabi (the chairman of the physics department, and winner of the Nobel prize for physics, 1944) and Willis Lamb, who was also to win a Nobel prize in 1955. Rabi exuded charisma and enthusiasm, but his lectures were confused and often incoherent. By contrast, Lamb's lectures were well organized and clear, but delivered in such a soft, droning voice that they tended to induce sleep. For years after we became colleagues, I would yawn compulsively on meeting Lamb. In time, I came to appreciate the depth and brilliance of Lamb's intelligence, but for a new graduate student who had not opened a book in more than three years, the effect was discouraging. I began to question my decision to switch from chemistry to physics.

About midway through my second year, I applied for a research position. Columbia was building a particle accelerator and they would need instruments and people who could be involved in the new 'high-energy physics'. This new subfield, which had emerged from nuclear and cosmic-ray physics, involved the study of a new class of subnuclear particle released by bombarding nuclei with high-energy particles from an accelerator. The postwar respect for the awesome weapons and devices that had been produced by physicists had provided generous government grants to advance research into the deepest levels of the atomic nucleus.

I was given the goal of building a 'Wilson cloud chamber', a device that could render visible the tracks of energetic particles made in the accelerator. No one at Columbia had any experience in this area, but Rabi — with wisdom and political cunning — had invited several experts to come to Columbia and help.

In 1948, I spent about a month away from the laboratory studying for my PhD qualifying exams. Exam-taking, especially for me, was traumatic and I looked forward to continuing to test the cloud-chamber device that we had designed from reading the literature.



Keeping in touch: Leon Lederman urges us to rediscover the day-to-day enjoyment of science.

Returning to the laboratory, I found a fellow mopping the floor and singing a fragment of Italian opera. As I entered, he snapped the mop in a military style and introduced himself: "Bernardini!" "Yes," I assured the new janitor (so I thought). "But be careful not to put water on all those wires." It took another 20 minutes to clarify that Gilberto Bernardini was not a new janitor, but a visiting professor from Rome with a very distinguished career in cosmic-ray physics. His English was primitive, mostly learned by reading the physics journals. For example, "Good evening" (it was morning), "It is evident that this is a nice day". More accurately, "Dis issa nice a day".

Bernardini had had a very difficult time in Italy during the war. Arriving in Columbia he was — if anything — as insecure as I was. But he soon started to change my insecurity and restore my original enthusiasms. He was 'in touch'. Through Enrico Fermi, his teacher, he knew the latest physics gossip and therefore he stimulated ideas on how to use the new accelerator. And the cloud chamber?

The principle was that one created all the conditions for rain inside a cylinder about 12 inches in diameter and 6 inches deep. However, the dense humidity created in this chamber requires some dust to trigger the formation of droplets. If the gas is too clean, no drops are formed. But an excited atom from which an electron was detached would serve. A train of droplets was the track of a particle that passed through the chamber. So we could 'see' subnuclear particles!

But my cloud chamber only gave a dense smoke. Bernardini looked at it. "Wazza dat?" He pointed to a long needle projected into the chamber. "That's my radioactive source." "Dakid oud!" he said. So I took it out, the

smoke disappeared and absolutely lovely tracks appeared. Bernardini started dancing.

Later, we made a proportional counter by machining a brass cylinder and stringing a wire through it, which exited through an insulating tube. A high voltage was applied while we flushed the cylinder with an appropriate gas. As we watched the results on an oscilloscope, pulses suddenly appeared.

"Izza counting!" Bernardini screamed. He lifted me (I was heavier than him) and danced me around. "What's happening?" I asked. "We seeing da particles from cosmic rays!"

He must have seen this thousands of times, but he never lost his sense of wonder. Some particle originating billions of miles away in some stellar catastrophe was being recorded in our cylinder on the tenth floor of the Columbia physics building! Fantastico!

Bernardini taught me the marvels of familiar phenomena. He would even turn on the lights in the laboratory, and turn them off and on again. How did this happen? What series of phenomena were organized to bring us light?

In my subsequent 40 years of research, there have been times of stress, frustration and disappointment. These are suffered in the hope that a discovery will bring everlasting joy, fame and fortune. But this is not a life. The fun and excitement must be daily — in the challenge of creating an instrument and seeing it work, the joy of communicating to colleagues and students, the pleasure of learning something new, in lectures, corridors and journals. And underlying it all, the sense of wonder that nature is comprehensible. The Italian visitor was my turning point. ■

Leon Lederman is director emeritus of Fermi National Accelerator Laboratory in Batavia, Illinois 60510-0500, USA.

Inside the oldest bird brain

Lawrence M. Witmer

Did *Archaeopteryx*, the most primitive known bird, have ‘the right stuff’? Looking into its skull with advanced technology provides insight into the dinosaurian transition to birds, and the evolution of flight.

Archaeopteryx may not be a global star of the calibre of, say, *Tyrannosaurus rex*, but it undoubtedly has iconic status. Combining the feathered wings and wishbone of birds with the toothed jaws and long bony tail of reptiles (Fig. 1), *Archaeopteryx* is the near-perfect transitional form. Since its discovery shortly after the publication of Charles Darwin's *Origin of Species* in 1859, it has been a compelling example in the case for evolution.

It has also been the central player in the debate on the origin of birds and avian flight. Although discoveries of feathered dinosaurs and archaic birds from China have advanced our understanding of the transition to birds^{1,2}, the *Archaeopteryx* skeletons collected from Jurassic limestones in southern Germany remain (at 147 million years old) the oldest undisputed avian fossils, and the most primitive³. The fossils have been scrutinized by so many scientists over the past 140 years that it might seem that nothing new could be learned. But a landmark study by Domínguez Alonso *et al.*, on page 666 of this issue⁴, goes back to the first skeleton ever found to present exciting data on the brain and sense organs. The results have implications for both the biology of *Archaeopteryx* and the evolutionary transition to birds.

Researchers at the Natural History Museum in London isolated the part of the skull that in life encased the brain (Fig. 2). The braincase is so tiny — smaller than the last segment of your little finger — that Angela Milner, the team leader, safely carried it in a box in her shirt pocket from London to the University of Texas at Austin, where it could be analysed with high-resolution X-ray computed tomography. Using X-rays, the team ‘sliced’ the braincase so finely (each slice less than half the thickness of a printed page of *Nature*) that they were able to peer inside the thin bone at the brain cavity and inner ear (the organ of balance and hearing), which was then digitally reconstructed.

Obtaining an understanding of the brain and sense organs is a top priority for palaeontologists, because such knowledge can offer insight into the behaviour of extinct organisms not otherwise provided by the skeleton. In the case of *Archaeopteryx*, from the beginning the question has been — could the oldest-known bird fly? In the past, answers have been sought from



Figure 1 Iconic *Archaeopteryx*. This is the ‘Berlin specimen’, discovered in 1877 and now kept at the Humboldt Museum in Berlin. The ‘London specimen’, which Domínguez Alonso *et al.*⁴ worked with, was discovered earlier (in 1861) and, although the head was separated from the body, it preserves the part of the skull enclosing the brain in exquisite detail.

aerodynamics, justifiably focusing on the structure of the wings and feathers^{3,5}. But flight isn’t just about wings, rudders and flaps. It’s also about the pilot and on-board computer, and those are the missing elements that this new study⁴ provides for *Archaeopteryx*.

The brain of *Archaeopteryx* was much like that of birds today, albeit of a primitive sort. It was larger than the brain of an average reptile of equivalent body size but smaller than any similarly sized modern bird brain. Its organization was also basically avian, with enhancement of those areas concerned with movement. Moreover, the visual centres are enlarged, suggesting that *Archaeopteryx* was a visually oriented animal. The new findings relating to the delicate inner-ear canals are particularly important, because recent studies have associated canal architecture with behaviour and mode of life^{6,7}. The canals of *Archaeopteryx* are again much more like those of birds than modern-day reptiles, suggesting that agility and coordination of head and eye movements were critical.

But is this the brain and ear of a flier? Some insight here can be provided by the entirely separate evolution of flight in pterosaurs, the extinct flying-reptile group

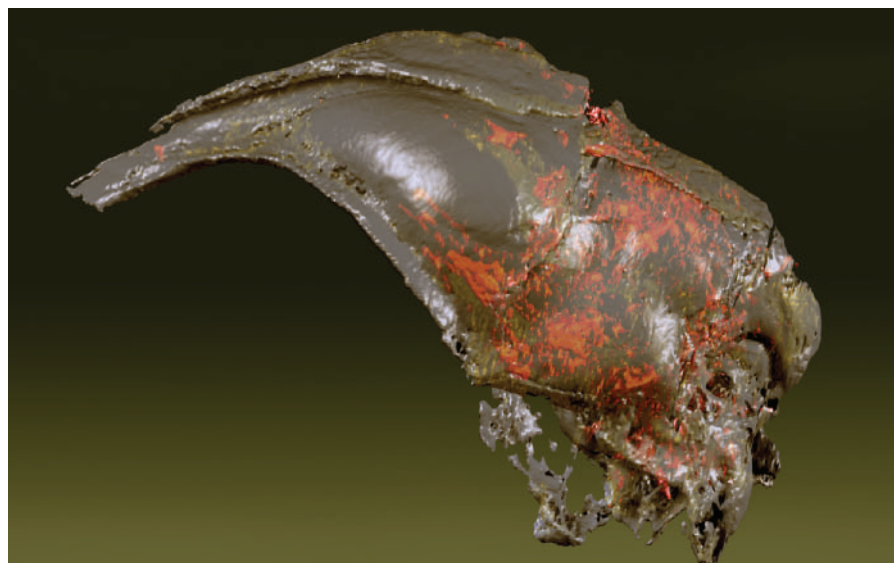


Figure 2 Bird brain? The three-dimensional reconstruction of the braincase and brain of *Archaeopteryx*, produced by Domínguez Alonso and colleagues using computed tomography. From their analyses of the brain and inner ear, they conclude that *Archaeopteryx* was probably equipped for flight. The reconstruction is about 20 mm in length; the red areas are crystals of manganese dioxide deposited during fossilization.

K. A. VON ZITTEL

P. DOMÍNGUEZ ALONSO

that dominated the skies between 230 million and 65 million years ago. A computed tomography study of pterosaurs⁶ that my colleagues and I carried out revealed expansion and reorganization of the brain and inner-ear canals very much like that seen in *Archaeopteryx*; and brain size relative to body size is almost identical. These independently evolved similarities suggest that there may indeed be some fundamental neural requirements for flight. In fact, Domínguez Alonso *et al.*⁴ argue that *Archaeopteryx* basically had 'the right stuff', from a neural standpoint, for flight. The authors also note, however, that some of the predatory (theropod) dinosaurs — the group that includes *Archaeopteryx* and all other birds — show some brain traits similar to those seen in birds. Domínguez Alonso *et al.* suggest that *Archaeopteryx* exhibits "a stage further towards the modern bird pattern". That hypothesis probably represents the most exciting outcome of this study: we finally have reliable data on the brain and inner ear of the most primitive known bird, and so can document the neural transition to birds.

Researchers will now race to the fossils of other early birds and bird-like theropods to look for the features identified in *Archaeopteryx*. What are the details of the transition? Did all of the 'avian' neural components evolve together or was this a piecemeal process? It might turn out that the non-flying progenitors of birds had developed many of these components. If so, then whereas pterosaurs clearly built their neural flight-control system from scratch⁶, birds may have evolutionarily co-opted for flight the advanced neural machinery they inherited, which was subsequently honed as flight improved. Perhaps most controversially, if a 'flight brain' or 'flight ear' can ever be characterized, can it provide a test of the heretical notion that some of the most bird-like Cretaceous theropods (such as *Velociraptor*) are actually the secondarily flightless descendants of early, *Archaeopteryx*-like birds⁸?

This latest in a long line of papers on *Archaeopteryx* affirms the iconic status of this fossil. It shows yet again that, in large measure, it all begins and ends with *Archaeopteryx*.

Lawrence M. Witmer is in the Department of Biomedical Sciences, Ohio University College of Osteopathic Medicine, Athens, Ohio 45701, USA.
e-mail: witmerL@ohio.edu

1. Xu, X. *et al.* *Nature* **421**, 335–340 (2003).
2. Zhou, Z., Barrett, P. M. & Hilton, J. *Nature* **421**, 807–814 (2003).
3. Elzanowski, A. in *Mesozoic Birds* (eds Chiappe, L. M. & Witmer, L. M.) 129–159 (Univ. California Press, Berkeley, 2002).
4. Domínguez Alonso, P., Milner, A. C., Ketcham, R. A., Cookson, M. J. & Rowe, T. B. *Nature* **430**, 666–669 (2004).
5. Rayner, J. M. V. in *Biomechanics in Evolution* (eds Rayner, J. M. V. & Wootton, R. J.) 183–212 (Cambridge Univ. Press, 1991).
6. Witmer, L. M., Chatterjee, S., Franzosa, J. & Rowe, T. *Nature* **425**, 950–953 (2003).
7. Spoor, F. *et al.* *Nature* **417**, 163–166 (2002).
8. Paul, G. S. *Dinosaurs of the Air: The Evolution and Loss of Flight in Dinosaurs and Birds* (Johns Hopkins Univ. Press, Baltimore, 2002).

Semiconductor physics

The value of seeing nothing

Jochen Mannhart and Darrell G. Schlom

Adding atoms to a semiconductor can improve its electronic properties. In an oxide, taking atoms away can have a similar electronic effect — one that could, it seems, be exploited in device applications.

By 2007, the information age will have hit a fundamental roadblock. Without major changes in technology, the spectacular improvements in computer performance that we have enjoyed for decades will cease, because transistors based on silicon and silicon dioxide will no longer be able to keep up with Gordon Moore's famous law^{1,2} — that the number of transistors per unit area in an integrated circuit doubles every couple of years. But these limitations might be overcome if Si and SiO₂ were complemented in these devices by other materials. The candidates of choice are oxides, which are already assuming a vital role in semiconductor electronics. Now Muller *et al.*³ (page 657 of this issue) show that it is possible to control the electronic properties of these materials with the nanoscale precision necessary for the information industry.

Oxides offer a broad spectrum of properties — some are excellent insulators, others are superconductors. Some oxides have flippable electric or magnetic dipoles,

suggesting myriad device possibilities. Indeed, oxides such as hafnium dioxide are forecast to replace SiO₂ in the transistors of laptop computers within only three years¹. Another oxide known as 'Lustigem' — alias strontium titanate (SrTiO₃) — was a popular diamond substitute in the 1960s. If some of its oxygen atoms are removed, the glittering gem turns a deep blue (Fig. 1), and changes from insulating to conducting. This change in colour and conductivity is due to electrons that are left behind: because there is a difference in charge between an oxygen ion (O²⁻) and an oxygen atom, for each oxygen atom removed two electrons are added to the SrTiO₃ matrix. Oxygen vacancies thus function as electron-donating dopants — an effect commonly achieved in semiconductors by replacing some atoms with others that contain more or fewer electrons than the atoms for which they substitute. But can doping through vacancies be implemented and monitored in a controlled way on the atomic scale?

It seems so. Muller and colleagues³ have made an unexpected double breakthrough. With unrivalled precision, they have measured the quantity and location of oxygen vacancies in films consisting of layers of fully oxidized SrTiO₃ and of SrTiO_{3-δ}, in which some oxygen atoms are missing. Their first major advance is to have grown alternating layers of doped (δ ≠ 0) and undoped (δ = 0) SrTiO_{3-δ}, where a layer may be as thin as three unit cells. Analogous 'superlattices' are used in conventional semiconductor technology to enhance the lifetime of charge carriers⁴; in oxide superconductors, they are used to increase the supercurrent density⁵. Muller *et al.* grew their superlattices using pulsed laser ablation — a popular research technique for depositing thin films of oxide materials. Deposition occurs when a laser beam hits a SrTiO₃ target inside a vacuum chamber, vaporizing its surface into a plasma. Some of the vaporized atoms condense on a nearby substrate, again of SrTiO₃, heated to 750 °C. Adjusting the oxygen pressure in the chamber controls the δ of the single crystalline SrTiO_{3-δ} layers deposited.

To image the oxygen vacancies, the authors used a scanning transmission electron microscope (STEM). As the tightly focused electron beam of the STEM is scanned across a cross-sectional slice of the deposited superlattice, a map is made of the positions where electrons are scattered slightly by oxygen vacancies and related



Figure 1 Now you see it, now you don't. These micrographs of a SrTiO₃ crystal show the effect of removing oxygen atoms, leaving vacancies in the crystal lattice: the glistening oxidized gem (top) is transformed into a dull blue, conductive crystal (bottom).

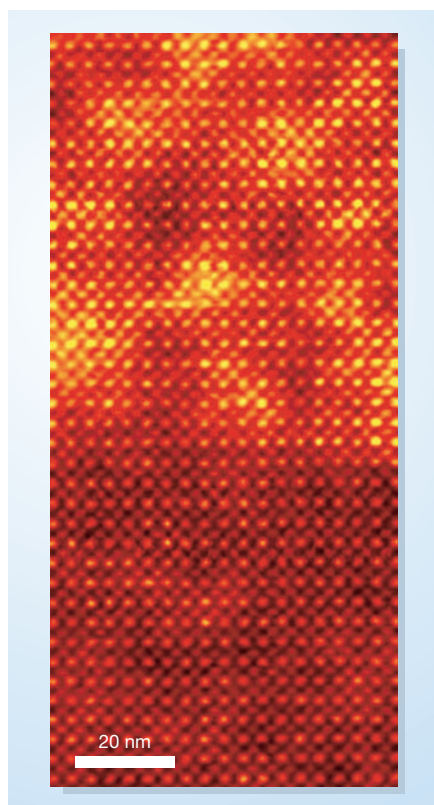


Figure 2 What a difference a layer makes. The abrupt junction between layers of SrTiO_3 (bottom) and $\text{SrTiO}_{3-\delta}$ (top) is clear in this image created by Muller *et al.*³ using a scanning transmission electron microscope. Each bright-orange blob is a cluster of oxygen vacancies.

defects; simultaneously, the energy loss of the transmitted electrons is measured, revealing the electronic effects of the missing oxygen atoms on the surrounding atoms (that is, changes in their oxidation state). This powerful technique⁶ offers outstanding sensitivity in resolving and identifying columns of atoms in crystalline samples, and has been used to image individual impurity atoms in silicon⁷.

The team scanned their layered samples in cross-section and spotted regions in which as few as two oxygen atoms were missing. And here came the second breakthrough. The STEM images (such as the one in Fig. 2) show that the oxygen vacancy concentration can change with surprising abruptness — from a layer with no oxygen vacancies to a layer with some constant number of vacancies over a distance of only 0.4 nm (the thickness of a single unit cell of SrTiO_3). At 700 °C, oxygen diffuses in minutes over many micrometres⁸, which would be expected to completely level out any nanometre-scale steps in the oxygen concentration profile. But it does not. That such sharp doping profiles are achievable is excellent news for the development of devices involving doped SrTiO_3 layers, as it has the highest mobility of any known oxide at low temperature⁹. Yet the data do raise the question of why the profiles are so crisp.

Are, for example, the oxygen vacancies or the sample microstructure stabilized by an as-yet-unknown mechanism, which may even be applicable to other ionic materials? No doubt Muller and colleagues will set about unravelling this puzzle too.

This work greatly broadens the options available for manipulating the electronic properties of oxides, and probably ionic materials of all sorts, at the nanometre scale. At present, the standard means of doping is to replace one cation with a different cation (that is, an impurity). But the ability to dope films without introducing impurities — thereby avoiding the risk that they might ride on the growing surface or hang around in the deposition chamber and become incorporated at undesired locations — is an enticing advantage of doping with vacancies. As films can grow by the lateral movement of steps as atoms are incorporated, even lateral doping using vacancies might be possible, analogous to the lateral superlattices and one-dimensional wires created using conventional

semiconductors¹⁰. Muller and colleagues show how a view of nothing can turn gems into electronics.

Jochen Mannhart is at the Centre for Electronic Correlations and Magnetism, Institute of Physics, University of Augsburg, D-86135 Augsburg, Germany. e-mail: jochen.mannhart@physik.uni-augsburg.de
Darrell G. Schlom is in the Department of Materials Science and Engineering, Penn State University, University Park, Pennsylvania 16802-5005, USA. e-mail: schlom@ems.psu.edu

1. International Technology Roadmap for Semiconductors, 2003 Edn, Front End Processes, 2 & 23–33 (Semiconductor Ind. Assoc., San Jose, 2003); <http://public.itrs.net/Files/2003ITRS/Home2003.htm>
2. Moore, G. E. *Electronics* **38**, 114–117 (1965).
3. Muller, D. A., Nakagawa, N., Ohtomo, A., Grazul, J. L. & Hwang, H. Y. *Nature* **430**, 657–660 (2004).
4. Tsu, R. & Esaki, L. *Appl. Phys. Lett.* **22**, 562–564 (1973).
5. Hammerl, G. *et al.* *Nature* **407**, 162–164 (2000).
6. Brown, L. M. *Nature* **366**, 721 (1993).
7. Voyles, P. M., Muller, D. A., Grazul, J. L., Citrin, P. H. & Gossmann, H.-J. L. *Nature* **416**, 826–829 (2002).
8. De Souza, R. A., Fleig, J., Merkle, R. & Maier, J. Z. *Metallk.* **94**, 218–225 (2003).
9. Tufte, O. N. & Chapman, P. W. *Phys. Rev.* **155**, 796–802 (1967).
10. Gossard, A. C. & Fafard, S. *Solid State Commun.* **92**, 63–70 (1994).

Fisheries science

Why mothers matter

Stephen R. Palumbi

Fish population growth depends on older mothers, which in some species produce more and 'better' offspring than younger fish. When fisheries remove the most productive females, the whole population suffers.

English literature about the ocean is predominantly male. Ernest Hemingway's *The Old Man and the Sea* was about an aggressive fight against the elements and the imponderable deep, with the main protagonist being the masculine *el mar*. But the real state of fisheries depends more on the role of females in the replenishment of fish populations. As Steven Berkeley and colleagues¹ now report in *Ecology*, that role has surprising facets. They find, contrary to popular wisdom, that in the black rockfish, older, larger female fish produce eggs and larvae that are much more likely to survive. From the standpoint of population growth rate and the potential to recover from overfishing, the old saying and country-and-western lyric may apply more often than Hemingway: "If Momma ain't happy, ain't nobody happy."

Larger female fish are vastly more productive than their smaller sisters. A single 61-cm-long red snapper (*Lutjanus campechanus*) has been estimated to produce as many eggs as 212 43-cm-long snappers². This is largely because eggs are produced in proportion to a fish's volume, which is proportional to the cube of its length. The prodigious fecundity of larger females has long been cited as a good reason to preserve fish populations in 'no-take' marine reserves. In some cases, larger fish in these reserves may

double a species' egg production — even if the reserves encompass only 5% of the marine habitat³.

Berkeley and co-workers¹ have documented another benefit from larger, older females. They studied rockfish of the genus *Sebastes* (Fig. 1, overleaf), and found that eggs from older females produced larvae that grew faster and were more resistant to starvation than larvae from younger females. The differences were huge: on the same diet, larvae of 12-year-old rockfish grew four times faster than larvae produced by 5-year-old rockfish. At the same time, offspring of older females had more metabolic reserves: larvae took an average of 12 days to starve whereas offspring of 5-year-olds starved in less than half that time.

The central difference lies in a small post-hatching gift each mother gives her offspring, a little oil droplet that serves as a metabolic reserve after the yolk-sac has been absorbed (Fig. 2). Older females provide a larger droplet than younger ones, ensuring a better head start for their larvae as they drift through the oceans, feeding and developing into juvenile fish capable of settling to the sea floor. Larger females, and females in better physical condition, produce better larvae as well, but these effects are slight compared with the effects of age. Such observations are

particularly surprising in light of long-held views that the optimum reproductive allocation per offspring should be independent of age or parent size⁴. Berkeley and colleagues are not sure why age makes such a difference, but there may be hidden age-related shifts in reproductive effort that will eventually provide an explanation.

These data are also important for attempts to rebuild overfished rockfish populations. Fishers value larger, older fishes — remember Hemingway — and strip away these larger individuals from the reproductive populations. In other fish, such as grouper species that change sex from smaller females to larger males, this tendency to take larger fish has long been known to reduce the number of mature males. Some grouper spawning aggregations have fewer than one male for dozens of females⁵, and this imbalance exhausts the fertilization abilities of the few surviving males. The opposite problem is seen in shrimp and crab populations where small males change into larger females⁶. Fishing down the family tree in these cases removes females first, and cuts egg production dramatically⁷.

The new data show why heavy fishing pressure on older fish is also a serious



Figure 2 Post-hatching gift. A larva from a 17-year-old black rockfish (*Sebastes melanops*), showing the large oil droplet used to fuel growth and stave off starvation. Younger mothers produce larvae with smaller droplets that are less likely to survive to the juvenile stage. (Courtesy S. Berkeley.)

problem in species that don't change sex. For example, in coastal Oregon, Berkeley and colleagues¹ document a decline in the average age of female rockfish from 9.5 years to 6.5 years during a period of intensified bottom fishing from 1996 to 1999. Such a culling of the older females reduces the average growth rate of larvae in the population by about 50%, and probably reduces the ability of these larvae to grow and survive to the next stage in their life history. Data on cod and haddock also show that larger females

produce larger eggs: presumably these larger eggs produce better larvae^{8,9}. The conclusion is that standard fisheries management tools that consider every female to be reproductively equivalent can be far off the mark.

Marine reserves change the landscape of fishing regulations by protecting the entire local populations of fish and invertebrates. The resulting dramatic reduction in mortality has an immediate benefit in producing larger fish in reserves — an effect that has been seen in tropical and temperate settings, along reefs, kelp beds and in estuaries. So protecting larger, older females can be more efficient in producing larger numbers of higher-quality offspring. Not all fisheries suffer from the need for more offspring. But because reproduction is the fuel that keeps all fisheries alive from one generation to the next, enhancing the success of larvae is a key to a sustainable fisheries future.

And Hemingway's Old Man Santiago may have known the power of females in the sea all along. Younger fishermen thought of the sea as masculine: "But the Old Man always thought of her as feminine and as something that gave or withheld great favours"¹⁰. Those favours are written in the next generation of the sea's creatures, and these new research results tell us that it is the favours of the mothers that matter most. ■

Stephen R. Palumbi is in the Department of Biological Sciences, Stanford University, Hopkins Marine Station, Pacific Grove, California 93950, USA.

e-mail: spalumbi@stanford.edu

- Berkeley, S., Chapman, C. & Sogard, S. *Ecology* **85**, 1258–1264 (2004).
- Bohnsack, J. *Aust. J. Ecol.* **23**, 298–304 (1998).
- Fujita, R. M., Willingham, V. & Freitas, J. *A Review of the Performance of Some United States West Coast Marine Reserves* (Environmental Defense Fund, Oakland, 1998).
- Smith, C. C. & Fretwell, S. D. *Am. Nat.* **108**, 499–506 (1974).
- Coleman, F., Koenig, C. & Collins, L. *Environ. Biol. Fishes* **47**, 129–141 (1996).
- Hannah, R. & Jones, S. *Fishery Bull.* **89**, 41–51 (1991).
- Charnov, E. L., Gershall, D. W. & Robinson, J. G. *Science* **200**, 204–206 (1978).
- Marteinsdottir, G. & Steinarsson, A. *J. Fish Biol.* **52**, 1241–1258 (1998).
- Hislop, J. *J. Fish Biol.* **32**, 923–930 (1988).
- Volcovici, V. *BookRags Book Notes on Old Man and the Sea* Quote 4, www.bookrags.com/notes/oms/QUO.htm

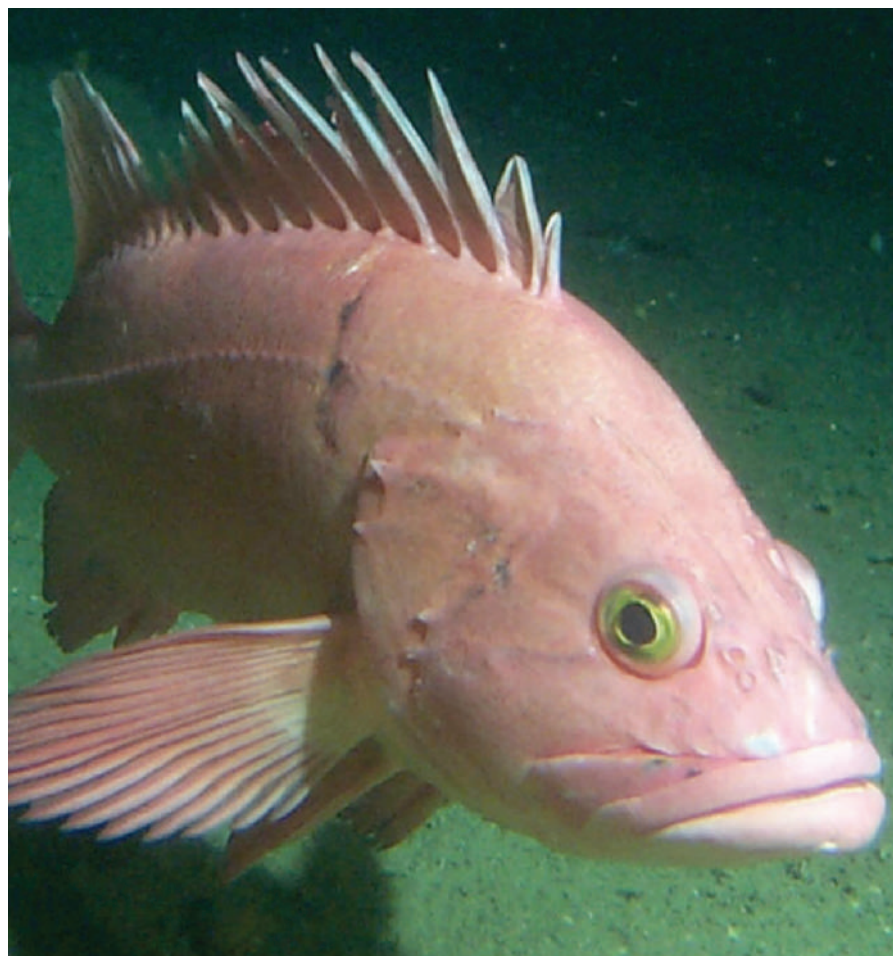


Figure 1 Yelloweye rockfish (*Sebastes ruberrimus*) from the west coast of North America. Berkeley *et al.*¹ find that older female rockfish produce larvae that have higher rates of growth and survival than those produced by younger mothers. (Courtesy Victoria O'Connell, Alaska Department of Fish and Game.)

Astronomy

A faint population of bursts?

Stan Woosley

An unusual γ -ray burst has been detected, much less energetic than expected. But its similarity to an earlier anomalous event suggests that lower-energy bursts might be more plentiful than had been thought.

It is seven years since astronomers were first able to pinpoint the origin of a γ -ray burst (GRB) in a distant galaxy^{1–3}. Since then, estimates of the energy associated with these titanic explosions have waxed and waned. Using only the brightness and distance of these objects, calculations had suggested that an amount of matter almost equal to the mass of the Sun was being transformed into γ -rays with near-perfect efficiency. But then evidence emerged for ‘beaming’ — the compression of the burst energy into narrow beams, or jets — and the energy estimates went down by a factor of about a hundred, to an amount equivalent to merely a few supernovae⁴. New evidence presented by Sazonov *et al.*⁵ and Soderberg *et al.*⁶ (starting on page 646 of this issue) suggests that the bursts observed so far are but the bright tip of the iceberg. It seems likely that many fainter explosions have gone unseen simply because our detectors weren’t sensitive enough.

The first indication that GRBs might have a broad range of energies came in 1998 with the discovery⁷ of a nearby event, ‘only’ 130 million light years away (equivalent to a redshift, z , of 0.0085). Despite its proximity, GRB 980425 — named for the day on which it was observed — was not particularly bright. Its inferred total energy, emitted over 20 seconds, was only 7×10^{47} erg — about 10,000 times less than a typical GRB, but still 10 trillion times more than the Sun emits in the same time. The burst was also ‘soft’ compared with other GRBs: no prompt emission was seen⁸ above an energy of 300 kiloelectronvolts (keV); and it was accompanied by a very bright supernova⁹, SN 1998bw, whose extreme luminosity, velocity, and radio emission placed it in a class by itself. Because of its low energy, however, astronomers argued as to whether GRB 980425 really was associated with the nearby supernova (from which the redshift was calculated), or whether it just happened to be in the same general direction but much farther away. Others argued that, even if the GRB and supernova were connected, the GRB was an unusual one, in a class by itself, with little relation to the others.

Time and further observation have diminished those arguments. X-ray observations of SN 1998bw spanning several years^{10,11} show a smooth evolution in brightness that matches that of the GRB, from just one day after its appearance. Moreover, as inferred from radio and X-ray emission, the

energy in the relativistic ejecta (the matter thrown out at more than 90% of the speed of light) was about 100 times more than that of the burst itself¹², and hence comparable to that of other GRBs. For some reason, GRB 980425 put a huge amount of energy into making a supernova (10^{52} erg), a lot of energy into relativistic ejecta (10^{50} erg), but relatively little into γ -rays moving towards Earth (10^{48} erg). Why was its γ -ray emission so low? And, given its faintness, might there be many other similar events that have escaped detection?

Then came GRB 031203. This burst was discovered from space last December by the γ -ray telescope INTEGRAL; the details of its γ -ray emission are now reported by Sazonov *et al.*⁵ and Soderberg *et al.*⁶. Like GRB 980425, the new burst lasted about 20 seconds and was close to our Galaxy. Most GRBs have a redshift that is greater than one, but the host galaxy of GRB 031203 was at $z = 0.1055$ (ref. 13) — making it, at 1.6 billion light years, the next-closest event after GRB 980425. From the distance and observed flux, the inferred energy of GRB 031203 was $0.6\text{--}1.4 \times 10^{50}$ erg, intermediate between GRB 980425 and more typical bursts. As the burst faded, a supernova was revealed^{14,15}, SN 2003lw, with properties similar to SN 1998bw. It looked as if a twin to GRB 980425 had been found.

But there were differences. Although faint compared with most bursts, GRB 031203 was a hundred times more luminous than GRB 980425; and its spectrum is ‘hard’, more like that of a typical GRB. A spectrum is described as hard if the typical energy of its γ -rays is high. Many scientists had come to accept (if not fully understand) an empirical relation¹⁶, called the Amati relation, between hardness and a burst’s total energy. But the hard spectrum of GRB 031203 promises to upset this apple-cart: the Amati relation would suggest a γ -ray energy of about 10 keV for GRB 031203; instead, INTEGRAL measured⁵ values in excess of 190 keV.

Another similarity between GRB 980425 and GRB 031203 was the energy they released in relativistic ejecta, inferred from their afterglows — the X-ray, radio and optical emission that follows the main burst for up to a few weeks, and now measured for GRB 031203 by Soderberg and colleagues⁶. For both GRBs, this energy was lower than average. Although the modelling of the process is uncertain, GRB 031203 may have been up to ten times less energetic⁶ than GRB

980425. The afterglow is created as the relativistic ejecta from the burst is slowed down by the surrounding medium. The emission from this more slowly moving matter is not nearly as concentrated into beams as the initial burst of γ -rays, so the energy in the afterglow is often regarded as a better measure of the total energy in the burst than the GRB itself (which is derived only from that part of the burst directed towards Earth). In fact, if the energies in relativistic ejecta and in the GRB itself are added together for GRB 980425 and for GRB 031203, these bursts are even more similar and distinctly less powerful than all other bursts analysed so far⁶.

So it seems that GRBs must now be considered a broad class of phenomena whose energy scale may vary by at least four orders of magnitude, even when they share such properties as spectrum, timescale, variability and the presence of a supernova. Perhaps this is not surprising. Models of GRBs offer considerable leeway in the partitioning of energy between supernova and relativistic ejecta, as it depends on such quantities as the rotation rate, mass and size of the parent star. However, given that most GRBs are beamed, their properties will inevitably vary depending upon the angle at which they are viewed. Sorting out both these effects — viewing angle and variable explosion energy — will require the study of many more bursts, especially the faint ones. Yet it is hard to detect events such as GRB 031203, which would have gone undiscovered if it had been just a few times more distant.

Fortunately, a new GRB mission, named Swift, will be launched by NASA this October. With its greater sensitivity, Swift should be able to detect many more faint bursts. Once it was thought that a more sensitive detector was not needed particularly urgently, because we could already see GRBs out over the entire observable Universe. But the data^{5,6} from GRB 031203 suggest that we may have been studying only the brightest ones. ■

Stan Woosley is in the Department of Astronomy and Astrophysics, University of California at Santa Cruz, Santa Cruz, California 95064, USA.
e-mail: woosley@ucolick.org

- Costa, E. *et al.* *Nature* **387**, 783–785 (1997).
- van Paradijs, J. *et al.* *Nature* **386**, 686–689 (1997).
- Metzger, M. R. *et al.* *Nature* **387**, 878–880 (1997).
- Frail, D. A. *et al.* *Astrophys. J.* **562**, L55–L58 (2001).
- Sazonov, S. Yu., Lutovinov, A. A. & Sunyaev, R. A. *Nature* **430**, 646–648 (2004).
- Soderberg, A. M. *et al.* *Nature* **430**, 648–650 (2004).
- Soffitta, P. *et al.* *IAU Circ. No.* 6884 (1998).
- Norris, J. P., Marani, G. F. & Bonnell, J. T. *Astrophys. J.* **534**, 248–257 (2000).
- Galama, T. J. *et al.* *Nature* **395**, 670–672 (1998).
- Kouveliotou, C. *et al.* *Astrophys. J.* **608**, 872–882 (2004).
- Pian, E. *et al.* *Adv. Space Res.* (in the press); preprint at <http://arXiv.org/abs/astro-ph/0304521> (2003).
- Li, Z. & Chevalier, R. A. *Astrophys. J.* **526**, 716–726 (1999).
- Prochaska, J. X. *et al.* *Astrophys. J.* (in the press); preprint at <http://arXiv.org/abs/astro-ph/0402085> (2004).
- Malesani, D. *et al.* *Astrophys. J.* **609**, L5–L8 (2004).
- Thomsen, B. *et al.* *Astron. Astrophys.* **419**, L21–L25 (2004).
- Amati, L. *et al.* *Astron. Astrophys.* **390**, 81–89 (2002).

Ultrafast physics

Quantum control with a twist

Yaron Silberberg

Laser pulses can be generated such that their shape and state of polarization change on the scale of a few femtoseconds, adding a new twist to the control and manipulation of molecules.

When driving a car, steering is crucial for getting where you want to go. Imagine driving a car with no steering-wheel in Manhattan: you could go to some places, along north–south avenues or east–west streets, but you would be very limited in your choice of destination. Laser scientists faced a similar limitation when trying to drive quantum systems, such as atoms or molecules, into desired states. The linearly polarized laser pulses they used could drive the system along one axis only; full control would require more complex light signals. But a series of advances in ultrafast optics — the latest reported by Suzuki *et al.*¹ and Brixner *et al.*² in *Physical Review Letters* — has now put a steering-wheel in their hands.

When a light pulse is short enough, it interacts with atoms and molecules before they can be affected by their environment. The interaction is then described by simple quantum mechanical rules. By carefully tailoring the shape of the optical pulse, it is possible to manoeuvre the system into desirable final states, particularly those that are hard to reach through simple thermodynamic processes. For example, it might be possible to break a certain bond in a molecule while leaving other, perhaps weaker,

bonds intact. The general approach is known as quantum coherent control, a field that developed as a theoretical exercise in the mid-1980s, but which has seen intense experimental effort in recent years³.

Coherent-control experiments start with light pulses that last typically a few tens of femtoseconds (one femtosecond is 10^{-15} seconds). Such pulses are now routinely produced by commercial lasers. The pulses are sent through an optical set-up known as a pulse shaper, which can be programmed to generate complex temporal shapes. The shaper acts as a frequency-domain synthesizer, separating a short pulse into many frequency components. The phase, and possibly the amplitude, of each component can be tweaked individually. The result is a longer pulse with an internal structure that can be defined with great precision.

It used to be the case that all quantum control experiments with shaped pulses used linearly polarized light — light whose electric-field vector is confined to a single direction. The optical field of a linearly polarized pulse puts a force on the charges in the system — be they electrons or ions — along only one direction. Now, it is quite easy to convert this linearly polarized light into

other polarization states, say circular or elliptical ones, by placing simple polarization converters in the beam. Yet this modifies the polarization of the entire pulse uniformly. Stretching the analogy made earlier, this is like driving a car whose steering-wheel is stuck, forcing it to turn constantly to one side — not a much better situation.

What is needed is the ability to change the driving direction continuously — that is, to modify the polarization direction within the optical field. This was recently realized by Brixner and Gerber⁴: in their polarization pulse shaper, not only the amplitude and phase but also the polarization state of the different frequency components can be changed, creating pulses with complex, twisted polarization structures. In such a pulse, the polarization direction may change on the scale of a few femtoseconds.

The polarization shaper was first used to perform a relatively simple task. By rotating the polarization of a narrow band of frequencies within the broader pulse spectrum, my group⁵ was able to generate two synchronized pulses whose polarization was at right angles to each other. This was used in a technique for nonlinear molecular spectroscopy known as CARS, which usually requires two laser sources. However, the main interest in polarization-shaped pulses stems from their ability to impose and control rotations. Light pulses with varying polarizations can transfer angular momentum and therefore either drive a system into a final rotating state or use rotational states as intermediates in more complex interactions. Both goals have been demonstrated in atomic and molecular systems.

In atoms, because of their symmetry, the situation is somewhat simpler than it is in molecules. Coherent control of the angular-momentum states of atoms using polarization-shaped pulses has been demonstrated by Dudovich *et al.*⁶: we showed that careful crafting of the polarization and phase of the pulse can excite, via multi-photon absorption, particular states that cannot be recognized by linearly polarized control. Controlling angular momentum in molecules is a greater experimental challenge. To start with, although molecules have preferred directions, in most experimental situations they are oriented randomly in space. In addition, the energy-level structure that is quite simple in atoms becomes more complex, as electronic states are combined with vibrational and rotational molecular states. This complexity demands more elaborate schemes for control that cannot rely on perfect knowledge of the system at hand.

Now the potential of polarization shaping for controlling molecular processes has at last been demonstrated. Suzuki *et al.*¹ and Brixner *et al.*² have investigated the ionization of diatomic molecules that were pre-aligned (to

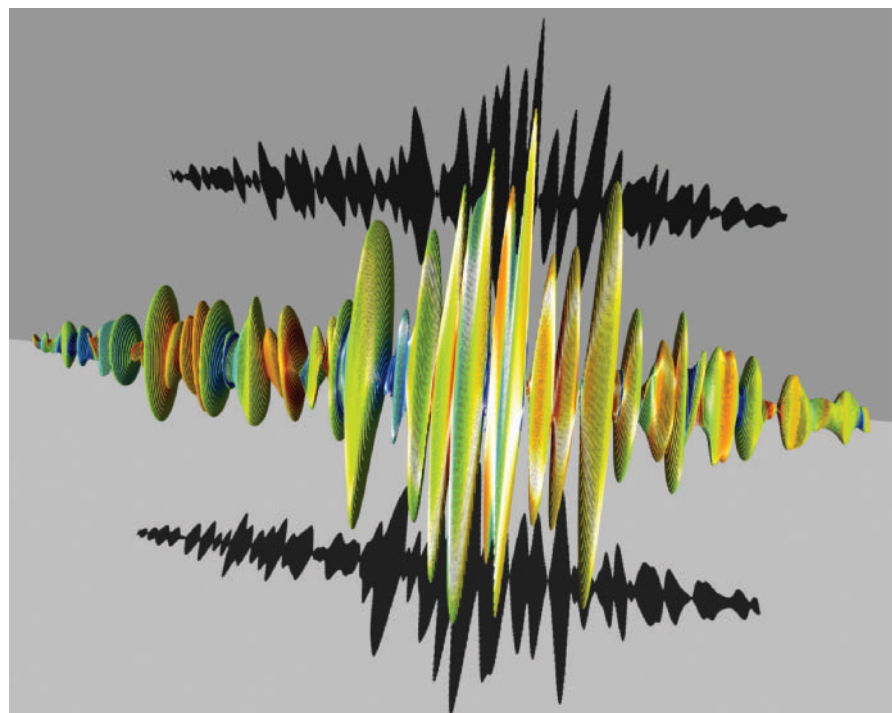


Figure 1 Polarization-shaped pulse, optimized for the ionization of potassium molecules². Ellipses represent the amplitude of the electric field; colours indicate different frequencies.

make sure they respond uniformly) and then irradiated with polarization-shaped pulses. These molecules — iodine¹ and potassium² — are as simple as molecules can be, but they are already too complex to theoretically design an optimized polarization structure that would maximize ionization. Therefore, both experiments relied on self-learning techniques, in which the optimal pulse shape and polarization structure (Fig. 1) are found through an iterative optimization procedure. Both groups have shown conclusively that pulses with complex polarization structures ionize these molecules more efficiently than pulses with a uniform polarization.

Further work will surely follow, using polarization-shaped pulses to tweak atomic and molecular systems with greater precision. Several proposals have already been

made and await experimental tests, including the alignment of molecules in a gas phase, the manipulation of chiral molecules and the control of attosecond pulse formation. Expect laser scientists to steer towards even more uses for these twisted pulses of light. ■

Yaron Silberberg is in the Department of Physics of Complex Systems, Weizmann Institute of Science, Rehovot 76100, Israel.

e-mail: yaron.silberberg@weizmann.ac.il

1. Suzuki, T., Minemoto, S., Kanai, T. & Sakai, H. *Phys. Rev. Lett.* **92**, 133005 (2004).
2. Brixner, T. *et al.* *Phys. Rev. Lett.* **92**, 208301 (2004).
3. Shapiro, M. & Brumer, P. *Principles of the Quantum Control of Molecular Processes* (Wiley, New York, 2003).
4. Brixner, T. & Gerber, G. *Opt. Lett.* **26**, 557–559 (2001).
5. Oron, D., Dudovich, N. & Silberberg, Y. *Phys. Rev. Lett.* **90**, 213902 (2003).
6. Dudovich, N., Oron, D. & Silberberg, Y. *Phys. Rev. Lett.* **92**, 103303 (2004).

Genomes

Worming into genetic instability

Susan M. Rosenberg and P. J. Hastings

A study of roundworms shows that genomic mutations occur surprisingly frequently, and that the kinds of changes involved differ from those predicted. Are genomes inherently less stable than previously suspected?

DNA carries the coded information that specifies the size, shape, body plan and many other basic characteristics of most organisms. To transmit these characteristics faithfully, DNA must pass from generation to generation with relatively few mutations. But mutations do happen, and can have profound consequences. These include inherited diseases, cancer and drug-resistant infections, but also the genetic differences among individuals that, through natural selection, drive evolution.

Until now, mutations seemed to be relatively rare and to occur in a characteristic spectrum. But such observations are challenged in the paper by Denver and colleagues on page 679 of this issue¹. These authors used a particularly powerful way to hunt for mutations in the roundworm *Caenorhabditis elegans* (Fig. 1, overleaf) — and found at least ten times more mutations, and a different assortment, than anticipated.

Traditional ways of estimating mutations are indirect², involving either phylogenetic studies of wild organisms or phenotypic methods in the lab. Phylogenetic studies involve comparing DNA sequences between species and estimating the number and kinds of changes that have occurred since the species diverged. Phenotypic methods rely on the ability of some mutations to change a trait (phenotype) of an organism. After a defined number of generations, rare mutants carrying the new trait are quantified, and mutation rates are calculated and then extrapolated to predict rates for the whole

genome. This extrapolation takes into account the genome's size and the fraction of mutations that has been estimated to produce phenotypic change (about one-third)².

However, both approaches probably underestimate the inherent mutation rate and skew the variety of mutations found. For instance, some mutations are harmful, and so the organisms that carry them are less likely to contribute to the next generation (they are 'selected against'), both in the wild and in large cultures. And the fraction of mutations that produces no phenotypic change might be larger than imagined.

Denver *et al.*¹ bypassed the phenotype-bias problem by directly sequencing randomly chosen stretches of DNA in laboratory-grown worms. They also minimized selection against harmful mutations by maintaining many lines of worms, separating a single worm from each progeny and allowing it to produce the next generation by self-fertilization, without competing with other worms. Rapid and severe loss of fitness occurs in these worms because, when their numbers are reduced to one repeatedly, random mutations become fixed — a phenomenon known as Muller's ratchet³.

From these pampered worms, Denver *et al.* sequenced four million base pairs of DNA, and found 30 new mutations compared with the original animals. This equates to a rate of 2.1 mutations per genome per generation. This rate is at least ten times higher than those reported previously in worms and other DNA-based organisms,



100 YEARS AGO

Prof. Schäfer, F.R.S., describes a simple and efficient method of performing artificial respiration in the human subject, especially in cases of drowning... Immediately the patient is recovered from the water he is placed face downwards, the head being turned sideways so that the mouth and nose are unobstructed, with a folded coat under the lower part of the chest; if respiration has ceased every instant of delay is serious. The operator then places himself athwart, or on one side of, the patient's body in a kneeling posture and facing the head. He places his hands flat over the lower part of the back (on the lowest ribs), one on each side, and gradually throws the weight of his body on to them so as to produce firm pressure — which must not be violent — on the patient's chest. This compresses the chest, and air (and water if there be any) is driven out of the patient's lungs. He then raises his body slowly so as to remove the pressure, still keeping his hands in position. This process of applying pressure and of relaxation of pressure by the forward and backward movement of the operator's body is repeated every four or five seconds without any marked pause between the movements. This course must be pursued for at least half an hour, or until the natural respirations are resumed... If there be means, others may remove the wet clothing by cutting it off, and may apply hot flannels to the body and limbs and hot bottles to the feet.

From *Nature* 4 August 1904.

50 YEARS AGO

The United Kingdom Atomic Energy Authority has announced that a heavy-water reactor (or atomic pile) which has been built at the Atomic Energy Research Establishment, Harwell, is now in operation. 'Dimple' is a low-powered thermal neutron research reactor. The heavy-water moderator is contained in a tank which is surrounded by a graphite neutron reflector. Outside this is a concrete radiation shield. The reactor fuel is submerged in the heavy water. Both the type of fuel and its arrangement in the tank can be changed quickly so that what is, in effect, a different design of reactor can be built up in a matter of days... The versatility of Dimple will make it an extremely valuable tool in the design of future power-producing reactors and for measuring essential constants in reactor physics.

From *Nature* 7 August 1954.

which curiously maintain constant predicted mutation rates per genome, irrespective of genome size². Moreover, the kinds of mutations differ from those previously seen in many organisms, even worms. Why so many and such different mutations?

Denver *et al.* suggest that the crucial difference lies in reducing the genetic selection that biases the accumulation of mutations. Specifically, they find that insertions of one to a few bases of DNA are more common than previously reported, and argue that these findings, when compared with the results of phylogenetic studies, suggest

that larger genomes are selected against in the wild. This might be true. But Denver *et al.* also found more insertions than are seen in lab-based studies², in which selection for small genome size should be minimal.

We suggest two other explanations. First, previous estimates of how many mutations give rise to phenotypic changes might be too low. Phenotypes might be masked by variables not factored in, such that mutation rates are higher than predicted from lab-based data — as Denver *et al.* observe. For example, phenotypes of altered proteins are known to be masked, or 'buffered', by molecular chaperones, molecules that spruce up distorted proteins into better approximations of their functional shapes. Loss of available chaperones — in response to 'heat shock', for instance — reveals many new phenotypes, produced by previous mutations, that are invisible with chaperones present⁴. Also, the numbers and kinds of mutations to which specific DNA sequences are prone might be an evolved trait⁵; perhaps most genes are actually less prone to mutations that cause phenotypic changes. So, correction factors for the number of genomic mutations per phenotypically detectable mutation may need to be revised upward. Denver and colleagues' data could help to provide an empirical correction factor.

A second, not mutually exclusive idea is that a state of increased mutability (a 'mutator' state) is induced in the worms specifically because they have accumulated harmful mutations through Muller's ratchet. We suggest that these mutations provoke cellular stress responses that, in turn, cause further mutations. The idea that Muller's ratchet can provoke general stress responses is supported by studies of the bacterium *Buchnera*, which endures severe population 'bottlenecks' and induces the production of a protein involved in the heat-shock/protein-stress response⁶. And in yeast, most



Figure 1 The roundworm — lessons in mutation.

gene-inactivating single mutations cause decreased fitness and altered patterns of expression of other genes⁷, perhaps reflecting the activation of stress responses.

Furthermore, at least two stress responses have been documented to cause mutations: the so-called SOS response to DNA damage in bacteria⁸, and the bacterial general-stress response, controlled by the RpoS protein^{9,10}, which increases the activity of some 50 genes in response to starvation and to temperature, pH, osmotic and oxidative stresses. One of these genes encodes a DNA-synthesizing enzyme¹¹ whose error-prone nature might underlie some of the mutation-inducing effects of this stress response^{9,10}. Moreover, as mentioned above, the heat-shock stress response exposes phenotypic variation⁴ — but might also cause genetic variation by increasing the mutation rate. We predict that the heat-shock response will indeed be found to cause mutations, and that many other general-stress responses will do so too.

It makes sense for stress responses to cause mutations; it may be a 'selected' feature that increases genetic variation, thus increasing 'evolvability' under stress when organisms are suboptimally adapted to their environments. Most of the mutations would be harmful or neutral, but rare adaptive mutations would also occur. Rare individuals in large stressed populations could thereby flourish, then, later, turn off their stress responses and readjust their mutation rates downwards to achieve greater genetic stability. Mutation-inducing stress responses might also underlie most cancers, which do not acquire mutator mutations, but still accumulate surprisingly high numbers of tumour-promoting alterations. Curiously, non-mutational gene-silencing events are common early in tumour progression, whereas mutations are more common later¹². Perhaps this is because early gene-

inactivating events cause mutation-inducing stress responses. Denver and colleagues' method and findings may provide windows on both of these important processes.

The nature of the alterations found by Denver *et al.*¹ supports the idea of a connection between mutations and stress responses. They strikingly resemble the mutations seen in cells lacking DNA 'mismatch repair'¹³, which corrects replication errors and is debilitated during a mutational stress response. Both the worms and mismatch-repair-defective cells show 10–100 times more mutations than normal

worms or cells; more small sequence insertions and deletions than base substitutions; more insertions and deletions in single-base repeats; and more 'transitions' than 'transversions' among base substitutions. We suspect that loss of mismatch-repair capacity is a general feature of mutation-promoting stress responses, as observed in starved bacteria¹⁴, in which a special error-prone DNA polymerase enzyme, induced by the SOS response⁸ and starvation¹¹, causes increased mutation¹⁵. In general, whenever excessive replication errors occur, mismatch repair should become exhausted^{14–16}. So Denver and colleagues' creatures may have wormed their way into genetic instability. Whatever the mechanism that accounts for their altered DNA, they are telling us something important about mutation. ■

Susan M. Rosenberg and P. J. Hastings are in the Department of Molecular and Human Genetics, and Susan M. Rosenberg is also in the Departments of Biochemistry and Molecular Biology, and Molecular Virology and Microbiology, Baylor College of Medicine, Houston, Texas 77030-3411, USA.

e-mail: smr@bcm.tmc.edu

- Denver, D. R., Morris, K., Lynch, M. & Thomas, W. K. *Nature* **430**, 679–682 (2004).
- Drake, J. W., Charlesworth, B., Charlesworth, D. & Crow, J. F. *Genetics* **148**, 1667–1686 (1998).
- Muller, H. J. *Mutat. Res.* **106**, 2–9 (1964).
- Rutherford, S. L. & Lindquist, S. *Nature* **396**, 336–342 (1998).
- Caporale, L. H. *Annu. Rev. Microbiol.* **57**, 467–485 (2003).
- Moran, N. A. *Proc. Natl Acad. Sci. USA* **93**, 2873–2878 (1996).
- Bergman, A. & Siegal, M. L. *Nature* **424**, 549–552 (2003).
- Sutton, M. D., Smith, B. T., Godoy, V. G. & Walker, G. C. *Annu. Rev. Genet.* **34**, 479–497 (2000).
- Lombardo, M. J., Aponyi, I. & Rosenberg, S. M. *Genetics* **166**, 669–680 (2004).
- Yang, H., Wolff, E., Kim, M., Diep, A. & Miller, J. H. *Mol. Microbiol.* **53**, 283–295 (2004).
- Layton, J. C. & Foster, P. L. *Mol. Microbiol.* **50**, 549–561 (2003).
- Suzuki, H. *et al. Nature Genet.* **36**, 417–422 (2004).
- Harfe, B. D. & Jinks-Robertson, S. *Annu. Rev. Genet.* **34**, 359–399 (2000).
- Harris, R. S. *et al. Genes Dev.* **11**, 2426–2437 (1997).
- McKenzie, G. J., Lee, P. L., Lombardo, M. J., Hastings, P. J. & Rosenberg, S. M. *Mol. Cell* **7**, 571–579 (2001).
- Schaaper, R. M. & Radman, M. *EMBO J.* **8**, 3511–3516 (1989).

Malaria

The genetics of resistance

Lancet **364**, 438–447 (2004)

In the borders of Thailand, drug-resistant forms of *Plasmodium falciparum* — the parasite that causes the most severe form of malaria in humans — are rife. Treatment for malaria in Thailand generally comprises the drugs mefloquine and artesunate, which are reasonably effective when given together. But resistance to this treatment regimen is a problem nevertheless, and resistance to mefloquine alone (the treatment of choice 20 years ago) is very common. The mechanisms of resistance are unclear, however.

Ric N. Price *et al.* show that the genetics of the parasite plays a part. Over a 12-year period, the researchers studied the responses of 618 malaria patients from Thailand to antimalarial drugs. Patients who were infected with parasites bearing extra copies of a particular gene, called *P. falciparum* multidrug-resistant gene-1 (*pfmdr1*), were most likely to be resistant to mefloquine when it was given alone or in combination with artesunate.

The authors hope that monitoring the number of copies of this gene in the parasites of malaria patients will aid epidemiological surveys of drug resistance in *P. falciparum*. In future, the technique might also be used to predict treatment failure in individual patients.

Helen Pilcher

Environmental chemistry

Blackout bonus

Geophys. Res. Lett. **31**, L13106 (2004)

The power cut of August 2003 may have caused misery for millions of people in the northeastern United States and southeastern Canada, but Lackson T. Marufu and colleagues jumped at the chance it offered to conduct a unique experiment in atmospheric chemistry. More than a hundred power plants were affected by the cuts, and the researchers recognized the unparalleled opportunity to assess the contribution the plants make to air pollution and smog.

The team collected air samples over the blackout area in Pennsylvania, analysing them in real time on board their aeroplane. They compared the results with data taken on the same day just outside the region, and with observations made the previous summer over the same location when meteorological conditions were similar and the power plants were fully operational.

A day into the blackout, ozone had decreased by nearly 50% and sulphur dioxide was down by 90%. The consequent reduction in sulphate particles, the principal



Inconvenient for some: New York in the August 2003 blackout.

cause of haze, meant that visibility improved by more than 40 km. Moreover, much of the eastern United States felt the benefit of the cleaner air, showing that emissions travel hundreds of kilometres from the power stations. The reductions are greater than expected from theoretical estimates, suggesting that models of emissions may have to be rethought.

Helen Dell

Stem cells

Ready for action

Cell **118**, 149–161 (2004)

Blood cells begin their lives as haematopoietic stem cells (HSCs) inside bone marrow. The bone marrow thus needs a population of these stem cells in a dormant, or quiescent, state, ready to proliferate and differentiate.

Fumio Arai *et al.* have now unravelled a signalling mechanism that keeps HSCs poised and preserved. Working on bone marrow from adult mice, the authors discovered that HSCs expressing the receptor protein Tie2 are quiescent, adhere to bone cells and do not die. What's more, the interaction of Tie2 with the angiopoietin-1 protein is sufficient to maintain the long-term blood-repopulating activity of HSCs *in vivo*.

The authors also point to the recently proposed concept of 'cancer stem cells' — in which cancer cells are replenished by

a mechanism rather like the one providing fresh blood cells. They suggest that signalling similar to that seen in HSCs may underpin the ability of leukaemia cells to maintain their populations in the bone marrow, thereby explaining how they can evade chemotherapy.

Michael Hopkin

Chemistry

Dyeing for cashew nuts

Ind. Eng. Chem. Res. doi:10.1021/ie030739s (2004)

Fuel oils are routinely dyed to ensure that they are not mixed, diluted or used for purposes other than that allowed under their tax rate. Diesel fuel for agricultural use is often sold at a much lower tax rate, for example, and is dyed to help government officials catch illegal users.

Somsaluay Suwanprasop *et al.* now report an ingenious new range of fuel dyes made from an agricultural by-product that is ubiquitous in Thailand: the shells of cashew nuts. The shells offer a cheap and plentiful source of cardanol, the starting material for an efficient synthesis of 15 different dyes. The dyes are normally invisible, but become coloured after a simple extraction procedure. They are also extremely soluble in fuel oils, and can be detected at just five parts per million — a concentration that has no discernible effect on the combustion properties of the fuel, unlike some other marker dyes.

Mark Peplow

Signal transduction

Flipping the switch

Mol. Cell **15**, 279–286 (2004)

Second messengers, such as cyclic AMP (cAMP) and cyclic GMP (cGMP), are small molecules that help to relay information from the outside to the inside of a cell. They influence a range of processes, including memory, inflammation and the immune response, and are controlled by a family of enzymes called phosphodiesterases. Now Kam Y. J. Zhang *et al.* have found a molecular switch inside phosphodiesterases that determines which of the second messengers interacts with the enzymes.

The researchers looked at the three-dimensional crystal structures of various phosphodiesterases, and found that the orientation of a key glutamine residue determines the enzymes' specificity. In one orientation, this amino acid interacts with cAMP; in another, it interacts with cGMP. In some enzymes, the glutamine residue can flip between orientations, allowing both second messengers to be controlled.

Many medicines, such as Viagra and the anti-inflammatory drug rolapram, target phosphodiesterases. Zhang *et al.* hope that their discovery will lead to the development of more selective and potent phosphodiesterase inhibitors.

Helen Pilcher

Surrogate broodstock produces salmonids

Trout offspring can be created from trout-donor germ cells transplanted into salmon.

A worldwide decline in the number of wild salmonids¹ calls for strategies to restore endangered populations. Here we show that germ cells can be transplanted between two different salmonid species, with the subsequent production of xenogenic, donor-derived offspring. This pioneering xenotransplantation technology may eventually find applications in facilitating the production of commercially valuable fish, as well as in species conservation.

Primordial germ cells (PGCs) are ideal donor cells because of their potential to develop into either male or female gametes². They have been successfully transplanted from one species into another in *Drosophila*³ and in birds⁴, but without the production of donor-derived offspring.

We have previously looked at PGCs in transgenic rainbow trout (*Oncorhynchus mykiss*) by incorporating a gene that encodes



Figure 2 Five-month-old trout juvenile produced from masu salmon surrogate sperm, derived from trout-donor germ cells.

green fluorescent protein (GFP)⁵; we have also used GFP-labelled PGCs as donor cells for transplantation between larvae of the same species⁶. To investigate whether the same technology can be applied between closely related species, with one acting as surrogate parents, we selected two salmonid species belonging to the genus *Oncorhynchus*, the rainbow trout and the masu salmon (*Oncorhynchus masou*), which were used as the donor and recipient, respectively.

The masu salmon is a Pacific salmon that is found only in east Asia, whereas the rainbow trout is native to North America. These two species have been phylogenetically separated for at least 8 million years⁷. The most striking biological difference between them is that rainbow trout are able to spawn several times during their lives, whereas masu salmon die after their first spawning.

We isolated 10–20 GFP-labelled PGCs from the genital ridges of newly hatched trout embryos and transplanted them into the peritoneal cavities of 60 newly hatched salmon embryos. At 30 days post-transplantation, the gonads of the recipient fish were examined for

colonization by donor PGCs (for methods, see supplementary information).

Donor-derived GFP-labelled germ cells were evident after 30 days in the gonads of 10 of the recipients ($16.7 \pm 4.4\%$), with an average count of 6.8 ± 1.8 donor-derived PGCs per recipient (Fig. 1a); we also found them in the testes of a male recipient after one year (results not shown) and in the ovary of a 17-month-old female recipient (Fig. 1b). These results confirm that trout PGCs are able to colonize salmon genital ridges and to differentiate into both male and female germ cells in the salmon gonads.

The presence of donor-derived spermatozoa in the milt of the male recipients was detected by using the polymerase chain reaction with GFP-specific primers (see supplementary information). Five milts from 37 male recipients were positive for the GFP gene (results not shown) and these were used to fertilize trout eggs artificially.

Progeny analysis indicated that trout offspring are produced when donor-derived trout spermatozoa (from the salmon) are used to fertilize trout eggs, whereas embryos are interspecific hybrids if trout eggs are fertilized by recipient-salmon spermatozoa. Of the first-generation offspring derived from the GFP-positive milt produced by one male recipient, 10 out of 2,332 (0.4%) embryos hatched at 34 days post-fertilization (early hatchlings), whereas the remainder hatched after 36–41 days (late hatchlings).

RAPD (for randomly amplified polymorphic DNA) analysis revealed that the DNA fingerprint of the early hatchlings was similar to that of the donor trout (Fig. 1c). In addition, the early hatchlings inherited the GFP transgene derived from the transplanted PGCs (Fig. 1d), indicating that the early hatchlings were produced from donor-derived trout sperm. The late hatchlings, on the other hand, were hybrids sired by recipient-derived salmon sperm (Fig. 1c). These hybrids had abnormal morphology and all died before reaching the 'swimming-up' stage, whereas the donor-derived offspring grew normally (Fig. 2). To our knowledge, this is the first time that it has been possible in any animal for xenogenic donor germ cells to develop into functional gametes that generate donor-derived individuals.

Male masu salmon generally reach sexual maturity a year after hatching, whereas rainbow trout males take two years to mature. As a result of our xenotransplantation technique, a one-year-old surrogate masu salmon male was able to produce functional rainbow-trout sperm. This technology would significantly reduce the time, cost, rearing space and

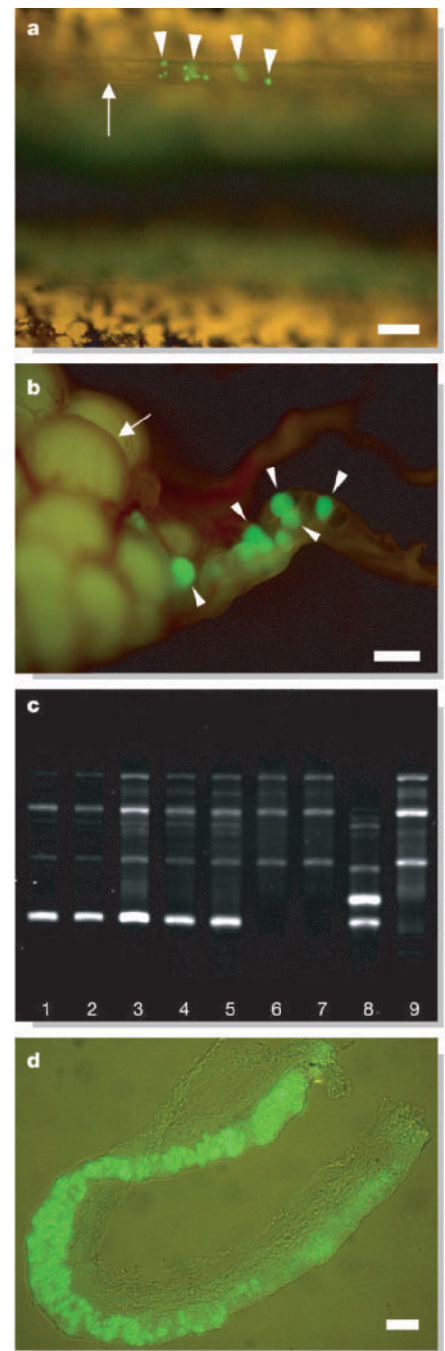


Figure 1 Colonization, proliferation and germline transmission of donor-derived trout germ cells transplanted into salmon embryos. **a**, Colonization by transplanted germ cells (arrowheads), labelled with green fluorescent protein (GFP), of a recipient's gonad (arrow) at 30 days post-transplantation. **b**, Donor-derived oocytes in a recipient-salmon ovary (arrowheads). Yolk-containing oocytes show yellowish-green autofluorescence (arrow). **c**, Randomly amplified polymorphic DNA analysis of late hatchlings (lanes 1–5) and early hatchlings (lanes 6, 7). Genomic DNAs from masu salmon (lane 8) and rainbow trout (lane 9) are included as controls. **d**, Excised genital ridge of an early hatchling, showing GFP-expressing germ cells. Scale bars: **a**, **d**, 100 μm ; **b**, 400 μm .

intensive labour that are normally necessary for the seed production of large donor species. Gamete production for a species that has a large body size and long generation time could be carried out in surrogate parents with a smaller body size and shorter generation time. As the cryopreservation of eggs and embryos has so far been unsuccessful for fish, PGC xenotransplantation, in combination with PGC cryopreservation⁸, could be a useful strategy for helping the conservation of endangered salmonid populations.

Yutaka Takeuchi*, Goro Yoshizaki*†, Toshio Takeuchi*

*Department of Marine Biosciences, Tokyo University of Marine Science and Technology, Minato-ku, Tokyo 108-8477, Japan
e-mail: goro@s.kaiyodai.ac.jp

†PRESTO, Japan Science and Technology Corporation, Kawaguchi-shi, Saitama 332-0012, Japan

1. Bella, D. A. *Science* **295**, 1236 (2002).
2. Wylie, C. *Cell* **96**, 165–174 (1999).
3. Lawrence, P. A., Ashburner, M. & Johnston, P. *Genetics* **134**, 1145–1148 (1993).
4. Ono, T., Yokoi, R. & Aoyama, H. *Exp. Anim.* **45**, 347–352 (1996).
5. Takeuchi, Y., Yoshizaki, G., Kobayashi, T. & Takeuchi, T. *Biol. Reprod.* **67**, 1087–1097 (2002).
6. Takeuchi, Y., Yoshizaki, G. & Takeuchi, T. *Biol. Reprod.* **69**, 1142–1149 (2003).
7. McKay, S. J., Devlin, R. H. & Smith, M. J. *Can. J. Fish. Aquat. Sci.* **53**, 1165–1176 (1996).
8. Kobayashi, T., Takeuchi, Y., Yoshizaki, G. & Takeuchi, T. *Fish Physiol. Biochem.* **23**, 479–480 (2003).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

Thin films

Unexpected magnetism in a dielectric oxide

It is generally accepted that magnetic order in an insulator requires the cation to have partially filled shells of *d* or *f* electrons. Here we show that thin films of hafnium dioxide (HfO₂), an insulating oxide better known as a dielectric layer for nanoscale electronic devices, can be ferromagnetic even without doping. This discovery challenges our understanding of magnetism in insulators, because neither Hf⁴⁺ nor O²⁻ are magnetic ions and the *d* and *f* shells of the Hf⁴⁺ ion are either empty or full.

The electronic spins of cations in insulating oxides are normally coupled by nearest-neighbour interactions — either superexchange or double exchange¹. No long-range magnetic order is anticipated below the percolation threshold, because there is no extended network of neighbouring sites occupied by magnetic ions. Controversy has therefore surrounded reports of high-temperature ferromagnetism in thin films of transparent, wide-bandgap oxides, such as zinc oxide, titanium dioxide or tin dioxide, that have been doped with a few per cent of a 3*d* cation such as V³⁺,

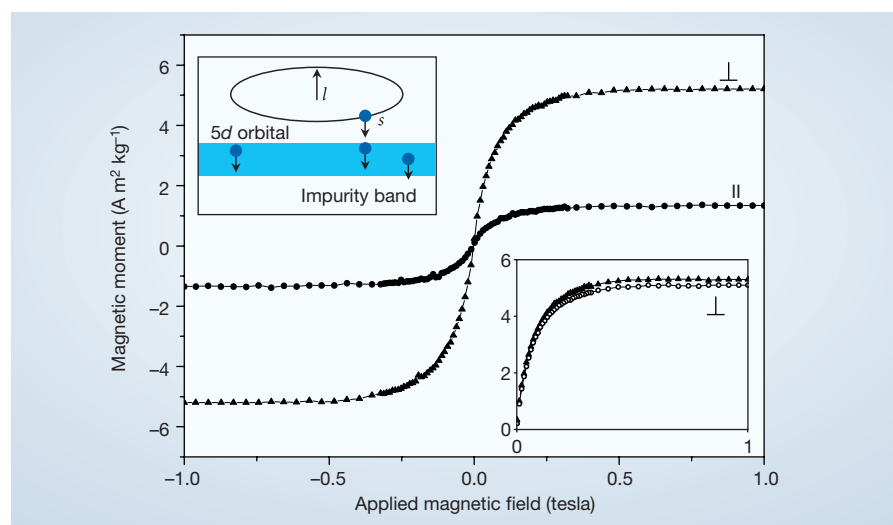


Figure 1 Magnetization curves for a thin film of hafnium dioxide measured in a superconducting quantum interference device (SQUID) magnetometer with the field parallel (||) or perpendicular (⊥) to the plane of an R-cut sapphire substrate. These room-temperature data are corrected for background diamagnetism of the substrate ($-1.8 \times 10^{-7} \text{ A m}^2 \text{ T}^{-1}$). Insets: top left, proposed coupling scheme for orbital, spin and impurity-band magnetism in hafnium dioxide (blue circles represent electrons, arrows show direction of their spin moment, the orbital and spin moments *l* and *s* of an electron in a 5*d* state couple antiparallel to form a *j* = 3/2 state); bottom right, magnetization curves at different temperatures (triangles, 5 K; circles, 400 K).

Mn²⁺, Fe³⁺ or Co²⁺ (ref. 2). The moments in these ferromagnets, which may be insulating or conducting, are sometimes too large to be explained by any impurity phase. An indirect exchange mechanism involving donor electrons in a spin-split impurity band has been proposed for such materials having a high dielectric constant (our unpublished results).

The Hf⁴⁺ ion has a closed shell [Xe]4*f*¹⁴ configuration, with no unpaired electrons. We deposited thin films of hafnium dioxide by conventional pulsed-laser deposition from a sintered target on to substrates maintained at 750 °C. The target was 99.95% pure, with zirconium as the main metallic impurity. Iron, cobalt and nickel were below 10⁻² wt%. Substrates were sapphire or silicon. The excimer laser wavelength and fluence were 248 nm and 1.8 J cm⁻², respectively. Oxygen pressure during deposition was varied from 10⁻⁴ to 1 millibar. Films were 45–135 nm thick, transparent, colourless and insulating; in a typical film, the *c*-planes of the monoclinic hafnium dioxide structure were oriented parallel to the surface of the R-cut sapphire substrate.

These films are ferromagnetic, with a Curie temperature exceeding 500 K and a magnetic moment of about 0.15 bohr magnetons per HfO₂ formula unit (Fig. 1). The magnetization is remarkably anisotropic, being up to three times greater when the magnetic field is applied perpendicularly to the plane of the film than when it is applied in the parallel direction. Coercivity at 5 K was about 5 millitesla. Results were reproduced in films of different thicknesses prepared at different oxygen pressures and on different substrates. No ferromagnetism was found in similarly prepared films of zinc or tin oxides.

These results confound our understanding of magnetism in solids. What in hafnium dioxide could possibly be magnetic? Lattice defects may be the key. In view of the preparation conditions, oxygen vacancies may be expected³, leading to *n*-type doping of the material. The electrons associated with defects in oxides having a high dielectric constant occupy large Bohr orbitals and tend to form an impurity band where they may be localized by correlations and local potential fluctuations⁴ to give the observed insulating behaviour. By allowing the impurity band to mix with the empty 5*d* states of hafnium and to transfer a fraction of an electron for each vacancy, the 5*d* states would in turn polarize the impurity band and provide the necessary ferromagnetic coupling. The anisotropy suggests a large orbital contribution to the 5*d* moment, with electrons in a spin-orbit coupled *j* = 3/2 state (Fig. 1, top-left inset).

Whatever the explanation, our discovery of anisotropic high-temperature *d*⁰ ferromagnetism in a transparent oxide is challenging for the theory of magnetism and propitious for the future of spin electronics, provided that the spin-polarized electrons can be mobilized.

M. Venkatesan, C. B. Fitzgerald, J. M. D. Coey

SFI-Trinity Nanoscience Laboratory, Physics Department, Trinity College, Dublin 2, Ireland
e-mail: jcoey@tcd.ie

1. Goodenough, J. B. *Magnetism and the Chemical Bond* (Interscience, New York, 1963).
2. Prellier, W., Fouchet, A. & Mercery, B. *J. Phys. Cond. Mat.* **15**, R1583–R1601 (2003).
3. Foster, A. S., Lopez-Gejo, F., Shluger, A. L. & Nieminen, R. M. *Phys. Rev. B* **65**, 174117 (2002).
4. Mott, N. F. *Conduction in Noncrystalline Materials* (Clarendon, Oxford, 1987).

Competing financial interests: declared none.

Pathways towards and away from Alzheimer's disease

Mark P. Mattson

Laboratory of Neurosciences, National Institute on Aging Intramural Research Program, 5600 Nathan Shock Drive, Baltimore, Maryland 21224, and Department of Neuroscience, Johns Hopkins University School of Medicine, 725 N. Wolfe Street, Baltimore, Maryland 21205, USA

Slowly but surely, Alzheimer's disease (AD) patients lose their memory and their cognitive abilities, and even their personalities may change dramatically. These changes are due to the progressive dysfunction and death of nerve cells that are responsible for the storage and processing of information. Although drugs can temporarily improve memory, at present there are no treatments that can stop or reverse the inexorable neurodegenerative process. But rapid progress towards understanding the cellular and molecular alterations that are responsible for the neuron's demise may soon help in developing effective preventative and therapeutic strategies.

Alzheimer's disease (AD) is a neurodegenerative disorder that currently affects nearly 2% of the population in industrialized countries; the risk of AD dramatically increases in individuals beyond the age of 70 and it is predicted that the incidence of AD will increase three-fold within the next 50 years (<http://www.alz.org>). Brain regions involved in learning and memory processes, including the temporal and frontal lobes, are reduced in size in AD patients as the result of degeneration of synapses and death of neurons (Fig. 1). Because there can be other causes of memory loss, definitive diagnosis of AD requires postmortem examination of the brain, which must contain sufficient numbers of "plaques" and "tangles" to qualify as affected by AD^{1,2}. Plaques are extracellular deposits of fibrils and amorphous aggregates of amyloid β -peptide ($A\beta$); diffuse deposits of $A\beta$ are also present in high amounts. Neurofibrillary tangles are intracellular fibrillar aggregates of the microtubule-associated protein tau that exhibit hyperphosphorylation and oxidative modifications. Plaques and tangles are present mainly in brain regions involved in learning and memory and emotional behaviours such as the entorhinal cortex, hippocampus, basal forebrain and amygdala. Brain regions with plaques typically exhibit reduced numbers of synapses, and neurites associated with the plaques are often damaged, which suggests that $A\beta$ damages synapses and neurites. Neurons that use glutamate or acetylcholine as neurotransmitters appear to be particularly affected, but neurons that produce serotonin and norepinephrine are also damaged. This review focuses on the molecular and cellular abnormalities that occur in the brain in AD, how they result in synaptic dysfunction and cell death, and how they can be counteracted. Central to the disease is altered proteolytic processing of the amyloid precursor protein (APP) resulting in the production and aggregation of neurotoxic forms of $A\beta$. Neurons that degenerate in AD exhibit increased oxidative damage, impaired energy metabolism and perturbed cellular calcium homeostasis; $A\beta$ appears to be an important instigator of these abnormalities. Genetic and environmental factors can determine one's risk for AD and an understanding of these risk factors and how they modify the amyloid cascade is leading to the development of interventions for the prevention and treatment of AD that range from changes in diet and lifestyle, to vaccines and drugs.

Pathways towards Alzheimer's disease

You cannot choose your parents

The genes passed on to you from your parents may or may not have a major role in determining whether you develop AD. The discoveries during the past 15 years of genetic aberrancies that either cause or

increase the risk of AD heralded a rapid increase in knowledge of the molecular and cellular alterations responsible for neuronal degeneration and cognitive dysfunction in AD³. Once the amino-acid sequence of $A\beta$ was determined, the gene encoding APP was cloned and localized to chromosome 21. DNA samples from pedigrees in which dominantly inherited AD occurred were then screened for mutations in APP, and several causative mutations were found⁴. Subsequent linkage analysis identified a region of chromosome 14 as the locus of a mutation(s) responsible for inherited AD

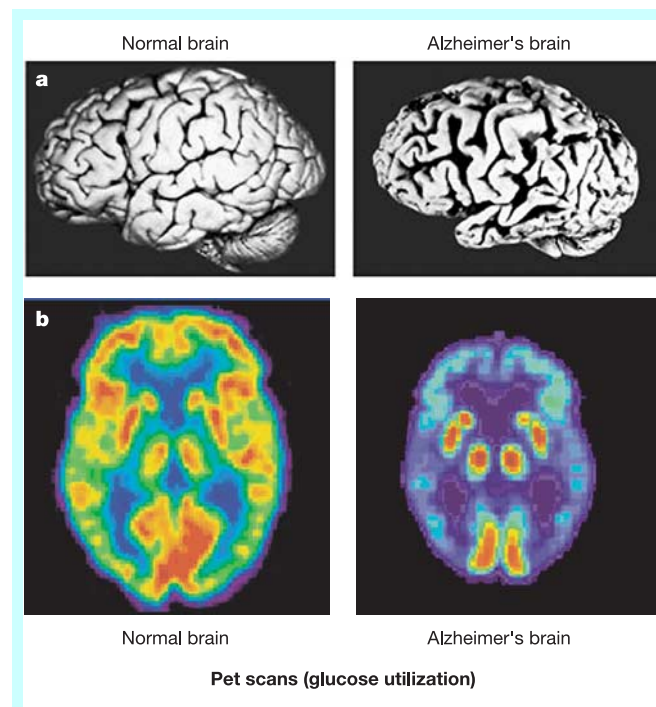


Figure 1 Alzheimer's disease results in shrinkage of brain regions involved in learning and memory which is correlated with major reductions in cellular energy metabolism in living patients. **a**, Compared with the brain of a healthy person, the brain of an Alzheimer's disease patient exhibits marked shrinkage of gyri in the temporal lobe (lower part of the brain) and frontal lobes (left part of the brain). **b**, Positron emission tomography (Pet) images showing glucose uptake (red and yellow indicate high levels of glucose uptake) in a living healthy person and a normal control subject. The Alzheimer's patient exhibits large decreases in energy metabolism in the frontal cortex (top of brain) and temporal lobes (sides of the brain).

Box 1

APP cut down to size

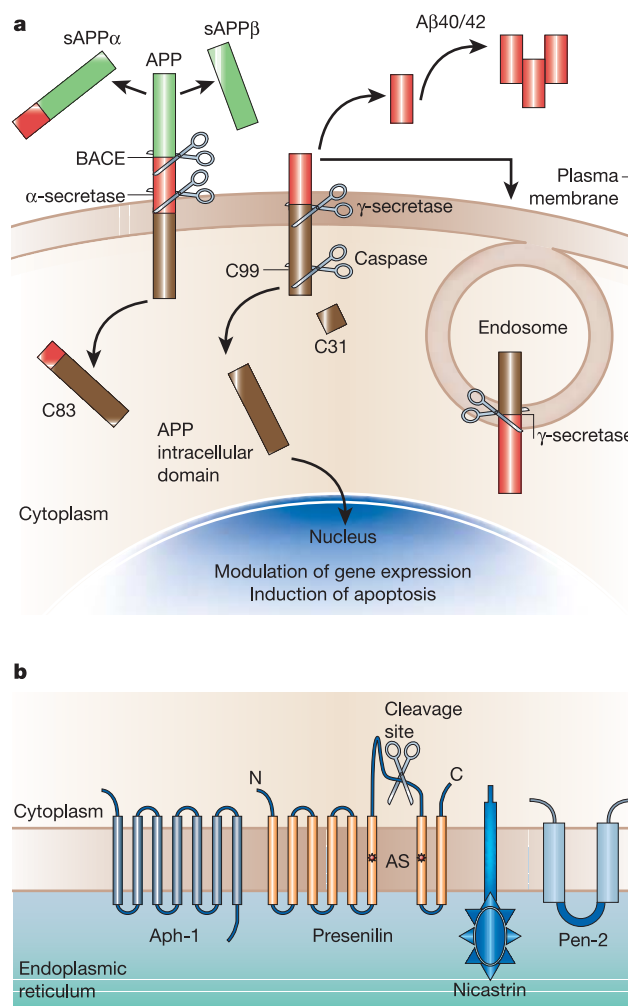
APP is widely expressed in cells throughout the body where the amount produced is influenced by the developmental and physiological state of the cells. APP is an integral membrane protein with a single membrane-spanning domain, a large extracellular glycosylated N terminus and a shorter cytoplasmic C terminus—A β is located at the cell surface (or on the luminal side of ER and Golgi membranes), with part of the peptide embedded in the membrane (Box 1 figure, panel a). APP is produced in several different isoforms ranging in size from 695 to 770 amino acids. The most abundant form in brain (APP695) is produced mainly by neurons, and differs from longer forms of APP in that it lacks a kunitz-type protease inhibitor sequence in its ectodomain^{4,17}. Enzyme activities involved in cleavage of APP at the α -, β - and γ -secretase sites are being identified (see ref. 3 for review). The identity of α -secretase remains unclear, although TACE (an enzyme responsible for cleavage of members of the TNF receptor family at the cell surface) and ADAM9 and ADAM10 are candidates^{64,65}. Cleavage of APP by α -secretase releases secreted (s)APP α from the cell surface and leaves an 83-amino-acid carboxy-terminal APP fragment (C83). Production of sAPP α increases in response to electrical activity and activation of muscarinic acetylcholine receptors, suggesting that neuronal activity increases α -secretase cleavage of APP¹⁷. Amyloidogenic processing of APP involves sequential cleavages by BACE and γ -secretase at the N and C termini of A β , respectively. The 99-amino-acid C-terminal fragment of APP generated by BACE cleavage can be internalized and further processed by γ -secretase to produce A β 40/42 in endocytic compartments. Cleavage of C99 by γ -secretase liberates an APP intracellular domain that can translocate to the nucleus where it may regulate gene expression, including the induction of apoptotic genes⁶⁶. Cleavage of APP/C99 by caspases produces a neurotoxic peptide (C31)⁶⁷.

γ -secretase, which cleaves APP within a transmembrane region, involves four different proteins, presenilin, nicastrin, Aph-1 and Pen-2 (Box 1 figure, panel b). The active site of γ -secretase requires the aspartyl protease activity of PS1 conferred by aspartate residues in adjacent transmembrane domains of the C- and N-terminal cleavage fragments of PS1 (red star). Nicastrin, Pen-2 and Aph-1 are each critical components of γ -secretase and each may modify enzyme activity in specific ways and in response to physiological stimuli^{68–70}. APP is only one of several proteins that is cleaved by γ -secretase, with Notch-1 being another well-studied γ -secretase substrate⁴. Notch-1 is a cell surface receptor that, when activated by ligands such as Jagged and Delta, is cleaved at the membrane resulting in the release of an intracellular domain of Notch. The intracellular domain of Notch then translocates to the nucleus, where it regulates the transcription of various genes. γ -secretase-mediated Notch-1 signalling plays an essential part in the regulation of cell fate during development of many organ systems including the brain as indicated by embryonic lethality and defective neurogenesis that is identical in Notch-1- and PS1-deficient mice⁷¹.

The normal functions of APP are not fully understood, but increasing evidence suggests it has important roles in regulating neuronal survival, neurite outgrowth, synaptic plasticity and cell adhesion¹⁷. APP is transported along axons to presynaptic terminals where it accumulates at relatively high levels, which can result in A β deposition at synapses⁷². One possible function of full-length APP is as a cell surface receptor that transduces signals within the cell in response to an extracellular ligand⁷³. However, neither a ligand nor downstream signalling cascades for APP have been clearly established. Physiological roles for sAPP α are supported by data showing that sAPP α is released from presynaptic terminals in response to electrical activity, and that sAPP α regulates neuronal excitability and enhances synaptic plasticity and learning and memory, possibly by activating a cell surface receptor that

modulates the activity of potassium channels and also activates the transcription factor NF- κ B¹⁶.

The initial experiments showing that synthetic fragments of A β can kill cultured neurons⁷⁴ led to a series of studies that have revealed the chemical and cell biological bases for the synaptic dysfunction and death of neurons in AD. A β may be most toxic when it is in the form of soluble oligomers in the earliest stages of aggregation^{75,76}. Synapses may be particularly susceptible to the adverse effects of aggregating forms of A β , as is suggested by the ability of A β to impair synaptic ion and glucose transporters and by electrophysiological studies showing that A β impairs synaptic plasticity^{17,77}. A β may damage neurons by inducing oxidative stress and disrupting cellular calcium homeostasis¹⁷. Coincident with the increased production of A β in AD is a decrease in the amount of sAPP α produced, which may contribute to the demise of neurons because sAPP α is known to increase the resistance of neurons to oxidative and metabolic insults¹⁷. In APP mutant mice memory deficits become apparent relatively early in the process of A β deposition, consistent with the neurotoxic effects of A β occurring during the formation of oligomers of the peptide⁷⁸. Adding to the evidence that A β deposition is a pivotal event in AD is the remarkable finding that immunization of APP mutant mice with human A β 42 results in the removal of A β deposits from the brain⁴⁴, which may result in reversal of cognitive deficits¹⁸.



in several different pedigrees, and the presenilin-1 (PS1) gene was identified as the affected gene⁵. Mutations in a gene on chromosome 1 with high homology to PS1, now called presenilin-2 (PS2), were then shown to cause a few cases of inherited AD⁶.

The vast majority of cases of AD are sporadic—they do not run in families. Nevertheless, molecular genetic analyses suggest that there are likely to be many genes that influence one's susceptibility to AD. The first such susceptibility gene identified was apolipoprotein E for which there are three alleles that encode three different isoforms of apolipoprotein E (E2, E3 and E4). Individuals that produce the E4 isoform are at increased risk of AD⁷. The mechanism whereby E4 promotes AD is not established, but there is evidence that E4 enhances A β aggregation and reduces A β clearance. In addition, data suggest that E4 might increase the risk of AD by enhancing amyloidogenic processing of APP, promoting cerebrovascular pathology, increasing oxidative stress and impairing neuronal plasticity. A second susceptibility locus for late-onset AD has been localized to chromosome 10, but the gene responsible has not yet been established⁸.

Risk factors

As with other age-related diseases (cardiovascular disease, diabetes,

cancers, and so on), there are likely to be behavioural, dietary and other environmental factors that may affect the risk of AD. However, this area of research has not yet matured to a point where clear recommendations can be made. Epidemiological findings suggest that a low education level, history of head trauma, consumption of high-calorie, high-fat diets and a sedentary lifestyle may each increase the risk of AD^{9,10}. When rodents are maintained in a cognitively stimulating environment or on a dietary restriction regimen, neurons in their hippocampus are more resistant to death and neurogenesis (the production of new neurons from stem cells) is increased^{10–12}. Similarly, regular physical exercise enhances hippocampal synaptic plasticity and neurogenesis, and is neuroprotective¹³.

Specific dietary components may affect the risk of AD. Individuals with low dietary folate intakes are at increased risk of AD, as an apparent consequence of increased levels of homocysteine; studies of mouse models of AD have demonstrated adverse effects of low dietary folate levels and high homocysteine levels on the disease process¹⁰. Other dietary factors implicated as risk factors for AD include lipids and metals such as copper and iron^{9,14}. However, despite accumulating data suggesting that dietary factors may influence disease risk, a causal relationship between caloric intake, or any specific dietary component, and AD has not been established.

Cleaving APP and unleashing A β

A fundamental abnormality that plays a pivotal role in the dysfunction and death of neurons in AD is altered proteolytic processing of APP, resulting in increased production and accumulation in the brain of neurotoxic forms of A β . The evidence supporting the “amyloid hypothesis” of AD is extensive and has recently been reviewed⁴. In every case of autosomal dominant early-onset familial AD where the genetic abnormality has been identified (mutations in the APP, PS1 and PS2 genes), the defective gene causes an increase in the production of the long 42-amino-acid form of A β (A β 42) in patients, in cultured cells and in transgenic mice^{15,16}. Three different enzyme activities have been identified that determine whether and which form of A β is produced (Box 1). The production of A β requires sequential cleavages of APP by β -secretase (BACE) at the N terminus of A β and by γ -secretase at the C terminus of A β . Alternatively, an enzyme called α -secretase cleaves APP within the A β sequence, thereby precluding production of A β . Mutations in APP that cause familial AD result in one or two amino-acid changes within or immediately adjacent to A β that enhance its cleavage by BACE and γ -secretase, whereas presenilin mutations alter γ -secretase activity⁴. The causes of altered APP metabolism and A β deposition in sporadic cases of AD are not understood, but may include age-related increases in oxidative stress, impaired energy metabolism and perturbed cellular ion homeostasis.

In addition to the histopathological and genetic evidence supporting a central role of APP mistreatment in AD pathogenesis, experimental findings in cell culture and animal models have identified adverse consequences of altered APP metabolism consistent with its involvement in synaptic dysfunction and nerve cell death in AD (Box 1). The increased A β deposition that occurs in AD most probably contributes to the demise of neurons because A β can be directly toxic to neurons and also greatly increases their vulnerability to oxidative and metabolic stress, and excitotoxicity¹⁷. The ability of immunization-based therapies that remove A β from the brains of APP mutant mice to improve synaptic function provides further evidence that A β plays a pivotal role in AD and also suggests that the process might be reversible (refs 4 and 18; see also section ‘Pathways away from AD’ below).

Renegade radicals and power cuts

Increased oxidative stress (uncontrolled production of highly reactive oxygen radicals) and impaired cellular energy metabolism are features of many prominent age-related diseases, and AD is no

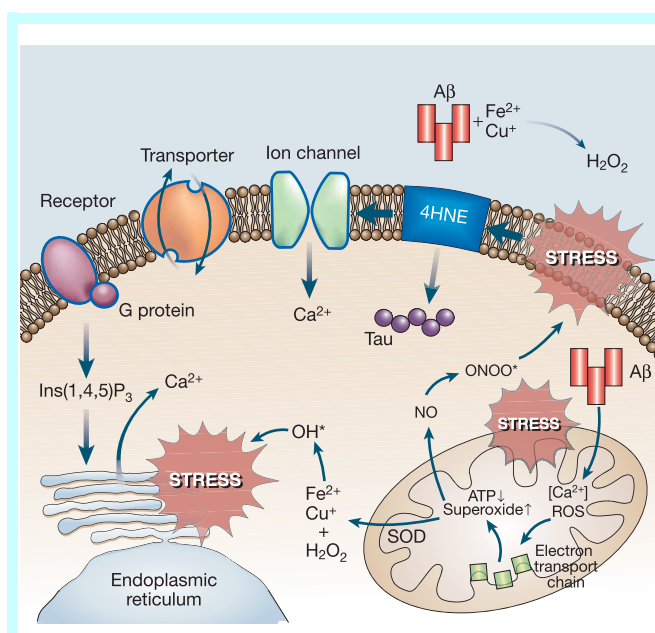


Figure 2 The neurotoxic action of A β involves generation of reactive oxygen species and disruption of cellular calcium homeostasis. Interactions of A β oligomers and Fe²⁺ or Cu⁺ generates H₂O₂. When A β aggregation occurs at the cell membrane, membrane-associated oxidative stress results in lipid peroxidation and the consequent generation of 4-hydroxynonenal (4HNE), a neurotoxic aldehyde that covalently modifies proteins on cysteine, lysine and histidine residues. Some of the proteins oxidatively modified by this A β -induced oxidative stress include membrane transporters (ion-motive ATPases, a glucose transporter and a glutamate transporter), receptors, GTP-binding proteins ('G proteins') and ion channels (VDCC, voltage-dependent chloride channel; NMDAR, *N*-methyl-D-aspartate receptor). Oxidative modifications of tau by 4HNE and other reactive oxygen species can promote its aggregation and may thereby induce the formation of neurofibrillary tangles. A β can also cause mitochondrial oxidative stress and dysregulation of Ca²⁺ homeostasis, resulting in impairment of the electron transport chain, increased production of superoxide anion radical and decreased production of ATP. Superoxide is converted to H₂O₂ by the activity of superoxide dismutases (SOD) and superoxide can also interact with nitric oxide (NO) via nitric oxide synthase (NOS) to produce peroxynitrite (ONOO[•]). Interaction of H₂O₂ with Fe²⁺ or Cu⁺ generates the hydroxyl radical (OH[•]), a highly reactive oxyradical and potent inducer of membrane-associated oxidative stress that contributes to the dysfunction of the ER.

Box 2

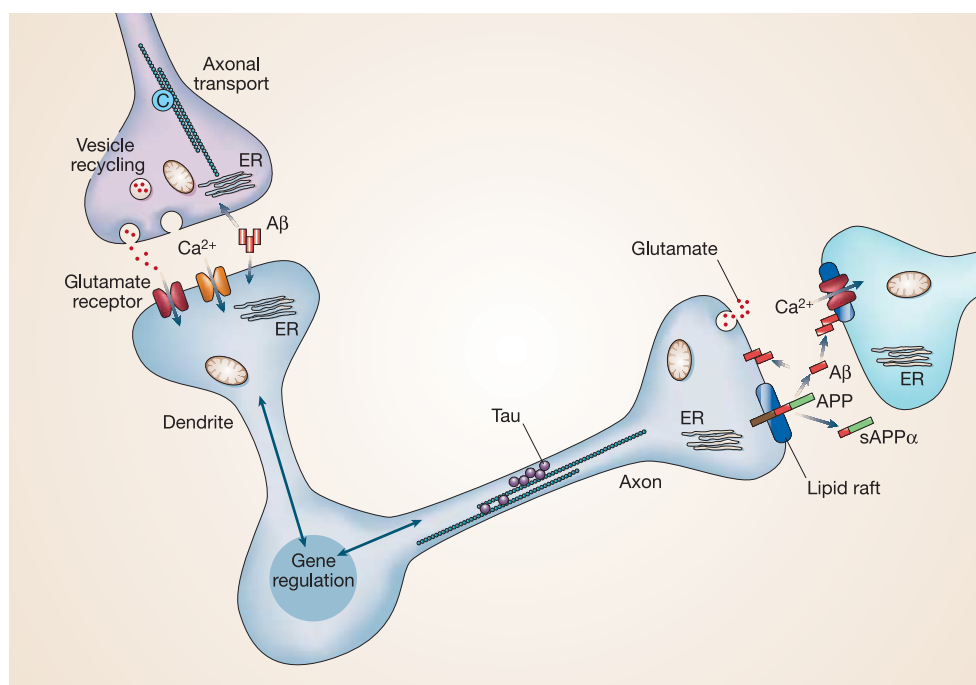
The weakest link?

Elaborate neuronal circuits with intricate morphologies and multiple synaptic communication sites allow the brain rapidly to process and store large amounts of information. Axons and dendrites that extend long distances from the cell body are exposed to many different environmental conditions and signalling molecules. To survive and function properly, synapses and axons possess local homeostatic mechanisms that protect them against adverse conditions, including those associated with ageing and disease. Several factors may render synapses and axons vulnerable in AD, including their high content of the disease-related proteins APP, presenilins and tau (axons), and their high metabolic and oxidative loads (Box 2 figure). Synaptic activity regulates APP processing and A β released from synaptic terminals may, in turn, modify synaptic plasticity⁷⁹. Because APP is axonally transported and processed in presynaptic terminals, synapses are sites where oligomers of A β may accumulate in high amounts⁸⁰. Soluble oligomers of A β 42 inhibit long-term potentiation (a memory-related form of synaptic plasticity) in the hippocampus of rodents, and increasing evidence suggests that such oligomers are responsible for memory impairment in AD patients⁸¹. In addition to A β , increased production of C-terminal fragments of APP may promote synaptic dysfunction and degeneration in AD⁸². In transgenic mouse models of AD, synaptic dysfunction and memory impairment can occur in the absence of any overt evidence of amyloid deposition or neuronal degeneration^{81,83,84}. Synapses are likely to be the sites at which neuronal death is initiated in AD because they contain most of the biochemical machinery for the initiation and execution of apoptosis, and A β can induce apoptotic cascades in synapses (ref. 85 and see Box 3).

Many of the neurons affected in AD are relatively large and have long axons. Analyses of AD patients suggest that damage to axons of such neurons may occur relatively early in the disease process⁸⁶. Alterations in axonal transport have been suggested to play a role in AD. APP is axonally transported to synaptic terminals in brain regions affected in AD, suggesting a role for axonal transport in plaque biogenesis⁸⁷. Other

findings suggest that abnormalities in tau may contribute to axonal transport deficits in AD⁸⁸, consistent with tau's function as an axonal protein that regulates microtubule stability. Perhaps the most compelling evidence that tau and axonal microtubule abnormalities play a fundamental role in the neurodegenerative process in AD comes from the causal role for tau mutations in familial frontotemporal lobe dementia with Parkinsonism, linked to chromosome 17 and elucidation of the pathogenic actions of these tau mutations in cell culture and animal models⁸⁹. Finally, alterations in the amounts and localization of synaptic proteins in vulnerable brain regions in AD patients have been documented⁹⁰, and may be involved in the early dysfunction of synapses. In particular, evidence is emerging that the process of synaptic vesicle recycling may be abnormal in AD⁹¹. Several proteins that regulate clathrin-mediated vesicle recycling are reduced in AD including synaptotagmin, dynamin, AP2 and AP180. Impairment of vesicle recycling might cause the enlargement of axon terminals documented in ultrastructural studies of AD and might also result in impaired neurotrophic factor signalling which involves clathrin-mediated endocytosis of activated receptors.

Perturbed processing of APP resulting in increased production of A β at synapses may be an early event in AD (Box 2 figure). APP is axonally transported and A β therefore likely accumulates at synapses in high amounts in AD. A β can have multiple adverse effects on the functions and integrity of both pre- and postsynaptic terminals including inducing oxidative stress, impairing calcium homeostasis and perturbing the functions of mitochondria and the ER. Abnormalities in axons may result from adverse effects of A β on tau, and microtubules resulting in neurofibrillary tangle formation and cell death. Presynaptic disturbances in synaptic vesicle trafficking and axonal transport may also contribute to the dysfunction and death of neurons in AD. Oxidative stress, perturbed calcium regulation and mitochondrial impairment are major alterations involved in functional and structural abnormalities in synapses and axons in AD.



exception. Cells in the brains of AD patients exhibit abnormally high amounts of oxidatively modified proteins, lipids and DNA; such free-radical-mediated molecular damage is particularly prominent in the environment of plaques and in neurofibrillary tangle-bearing neurons, suggesting roles for oxidative stress in amyloid-mediated neuronal damage and neurofibrillary pathologies¹⁹. Several sources of oxidative stress in AD have been proposed, with A β and redox-active metals such as Fe²⁺ and Cu⁺ being two such sources^{14,19,20}. During the process of aggregation A β generates hydrogen peroxide, a process that requires oxygen and that is greatly potentiated by Fe²⁺ and Cu⁺ (refs 13, 18). Lipid peroxidation induced by A β impairs the function of ion-motive ATPases, glucose and glutamate transporters, and also GTP-binding proteins as the result of covalent modification of the proteins by the aldehyde 4-hydroxynonenal¹⁷. By disturbing cellular ion homeostasis and energy metabolism, relatively low levels of membrane-associated

oxidative stress can render neurons vulnerable to excitotoxicity and apoptosis. The dysfunction and degeneration of synapses in AD may involve A β -induced oxidative stress, because exposure of synapses to A β impairs the function of membrane ion and glutamate transporters and compromises mitochondrial function by an oxidative-stress-mediated mechanism (Fig. 2).

Brain imaging studies have demonstrated deficits in glucose use in living AD patients, an abnormality that may occur before the onset of clinical symptoms²¹. The activities of cytochrome *c* oxidase, pyruvate dehydrogenase complex and α -ketoglutarate dehydrogenase complex—critical enzymes in energy metabolism—are decreased in brain cells of AD patients²¹. Altered proteolytic processing of APP may contribute to impaired energy metabolism, because brain glucose metabolism is decreased in cognition-related brain regions of APP mutant mice in association with increased amounts of A β in their brains^{13,22}. Moreover, hypoxic tolerance is significantly

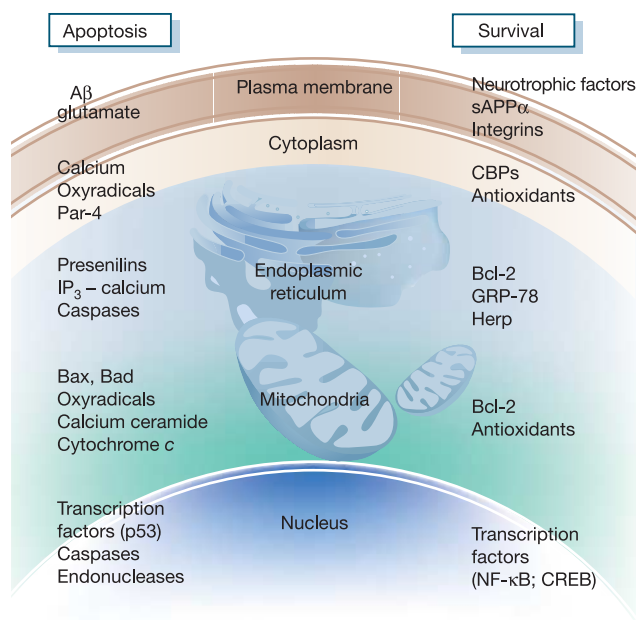
Box 3

The death of neurons

Apoptosis is a form of programmed cell death that involves changes in the cytoplasm, ER, mitochondria and nucleus (Box 3 figure). It typically includes the production and/or activation of proteins such as Bax and Bad that increase the permeability of mitochondrial and ER membranes resulting in the release into the cytoplasm of cytochrome *c* from the mitochondria and calcium from the ER⁹². The latter events then activate enzymes called caspases that cleave various protein substrates to sculpt morphological and biochemical aspects of the cell death process. For example, cleavage of cytoskeletal proteins and ion channel proteins causes cell shrinkage and may prevent necrosis, whereas cleavage of DNA degrades chromosomes. Evidence that many neurons undergo apoptosis in AD includes the presence of high levels of activated apoptotic proteins such as caspase-3 and Bax in neurons that exhibit neurofibrillary tangle pathology^{93,94}. In addition, DNA damage and upregulation of the pro-apoptotic proteins p53 and Bax occur in vulnerable neuronal populations at a relatively early stage in the disease process. Familial AD (FAD) mutations in presenilins render neurons vulnerable to apoptosis induced by A β , trophic factor deprivation and other stimuli⁹³, consistent with an apoptotic mode of neuronal death in patients with these presenilin mutations. APP mutations are also sufficient to trigger apoptosis in cultured cells⁹⁵. AIF, apoptosis-inducing factor; ATM, ataxia telangiectasia mutated; Apaf-1, apoptotic protease-activating factor 1. CREB, cyclic AMP binding protein; CBP, CREB binding protein; IP₃, inositol 1,4,5 trisphosphate

The triggers of cell death in AD may include A β , activation of glutamate receptors, increased oxidative stress, DNA damage and elevation of intracellular calcium levels. Once triggered, apoptosis proceeds with the production and/or activation of proteins such as p53, Bax, Bad and Par-4 that induce mitochondrial membrane permeability changes. Release of cytochrome *c* and AIF (apoptosis-inducing factor) from mitochondria and release of calcium from the ER occurs, followed by activation of the apoptosome, a protein complex that includes cytochrome *c*, Apaf-1 and caspases 9 and 3. Finally, cleavage of various substrate proteins by caspases and endonucleases occurs. Neurotoxic forms of A β may be a trigger of apoptosis in AD because pro-apoptotic proteins are associated with A β deposits in the brains of AD patients and A β can induce apoptosis of cultured neurons^{93,94}. Exposure of neurons to A β induces an apoptotic cascade that involves: upregulation of p53, Bax and Par-4; mitochondrial membrane permeability transition and release of cytochrome *c*; activation of the apoptosome resulting in caspase-3 activation; and nuclear chromatin fragmentation. Neurons treated with inhibitors of p53, agents that stabilize mitochondrial and ER membranes or caspase inhibitors are resistant to being killed by A β ^{93,94}. In addition to A β , C-terminal proteolytic products of APP have been implicated in neuronal apoptosis^{67,95}.

In addition to mitochondrial alterations, abnormalities in the ER and nucleus have been documented in studies of AD patient tissue and cell culture and animal models. FAD PS1 mutations have been shown to render neurons vulnerable to apoptosis, apparently by altering ER stress responses and ER calcium regulation^{29,84}. Apoptosis in AD may involve DNA-damage-response pathways and inappropriate activation of cell cycle pathways involving the cyclin-dependent kinase 5 and its neuron-specific activator p35. As evidence, proteolytic cleavage of p35 produces p25, which accumulates in the brains of AD patients and in cultured neurons exposed to A β 1–42 (ref. 96), and A β -induced neuronal death is mediated by the ATM-dependent DNA-damage-response pathway⁹⁷. Increasing evidence suggests that ubiquitin-proteasomal degradation of proteins is impaired in AD, resulting in the abnormal accumulation of damaged proteins in neurons (UBB + 1 mutations)⁹⁸. In this regard it was recently reported that a ubiquitin-conjugating enzyme called E2-25K/Hip-2 mediates A β -induced inhibition of proteasome activity, which is required for A β -induced apoptosis⁹⁹. There are several mechanisms that can protect neurons against apoptosis, including activation of cell surface receptors coupled to anti-apoptotic pathways, such as those activated by neurotrophic factors and sAPP α ¹⁶, and agonists that activate the cyclic AMP second messenger pathway^{17,100}.



decreased in presymptomatic APP mutant mice, suggesting an early role for perturbed energy metabolism in the pathogenic action of altered APP processing²³. There is increasing evidence that a systemic abnormality in glucose regulation, and insulin resistance in particular, is a risk factor for AD²⁴, although the possible link between diabetes and AD requires further investigation. The ability of dietary restriction, which enhances insulin sensitivity, to protect neurons in experimental models relevant to AD supports a role for perturbed glucose metabolism in AD pathogenesis¹⁰, although this remains to be established in humans. Because oxidative stress and impaired energy metabolism can induce amyloidogenic processing of APP, resulting in accumulation of potentially neurotoxic forms of A β , it has been suggested that such cellular stresses may promote amyloidogenesis²⁵. The latter mechanism could contribute to increased amyloid production in late-onset forms of AD, because oxidative stress and metabolic impairment increase with advancing age.

Uncontrollable calcium

The calcium ion (Ca²⁺) plays fundamental roles in learning and memory and is also involved in neuron survival and death. The inability of neurons to regulate calcium homeostasis is an aspect of AD pathogenesis that appears to be intimately involved in the dysfunction and death of neurons²⁵. Studies of AD patients and of the pathogenic actions of APP and PS1 mutations support a role for perturbed calcium regulation in AD. For example, neurofibrillary tangle-bearing neurons exhibit high amounts of Ca²⁺ and evidence of hyperactivation of Ca²⁺-dependent proteases and Ca²⁺-activated kinases²⁶. AD-causing mutations in PS1 and APP perturb cellular calcium homeostasis in cultured neurons and transgenic mice. Perturbed processing of APP may destabilize calcium homeostasis in neurons by increasing production of A β 42 and by decreasing levels of sAPP α ²⁵. A β may perturb calcium regulation by inducing oxidative stress, which impairs membrane calcium pumps and enhances calcium influx through voltage-dependent channels and ionotropic glutamate receptors (Fig. 2). Other findings suggest that A β can promote Ca²⁺ influx by forming channels in membranes or by activating cell surface receptors coupled to calcium influx^{27,28}.

Presenilin mutations and amyloidogenic processing of APP can destabilize Ca²⁺ homeostasis in astrocytes, oligodendrocytes and microglia, which might contribute to white-matter damage and local inflammatory processes²⁷. AD-causing PS1 mutations have been shown to alter Ca²⁺ regulation by increasing the amount of Ca²⁺ that can be released from the endoplasmic reticulum (ER) in neurons stimulated by glutamate, membrane depolarization or agonists that activate receptors coupled to inositol-1,4,5-trisphosphate (Ins(1,4,5)P₃) production²⁹. Moreover, a defect in capacitative calcium entry has been documented in cells expressing mutant PS1. However, presenilin mutations are a very rare cause of AD, and there is no evidence that a Ca²⁺-regulating action of PS1 is involved in sporadic forms of AD. Disturbances of Ca²⁺ regulation in AD may not be limited to neurons. Nevertheless, recent findings from studies of neurons lacking PS1 suggest that wild-type PS1 may normally serve a Ca²⁺-regulating function in the ER³⁰. Interestingly, studies of lymphocytes from patients with familial or sporadic AD, and from APP and PS1 mutant mice, have demonstrated abnormalities in calcium signalling reminiscent of those in neurons³¹. Thus, perturbed cellular Ca²⁺ homeostasis appears to be a widespread abnormality in both familial and sporadic forms of AD that may contribute to the disease process.

Feeding in the fats

Lipids are structural components of cell membranes, serve as intra- and intercellular signalling molecules and can modify the functions of many different proteins. Individuals that consume diets high in cholesterol and those with increased cholesterol levels may be at increased risk of AD, whereas those who take cholesterol-lowering drugs (statins) may be at reduced risk^{32,33}. The well-known adverse effects of high-cholesterol diets on the vasculature could contribute to an increased risk of AD. However, accumulating data suggest that cholesterol may feed more directly into the amyloid cascade by promoting amyloidogenic processing of APP. Statin treatment decreases A β levels and plaque formation in APP mutant transgenic mice³⁴, and high levels of cholesterol and shifts in subcellular cholesterol metabolism can increase the production of A β in cultured cells and APP mutant transgenic mice³². Alterations in

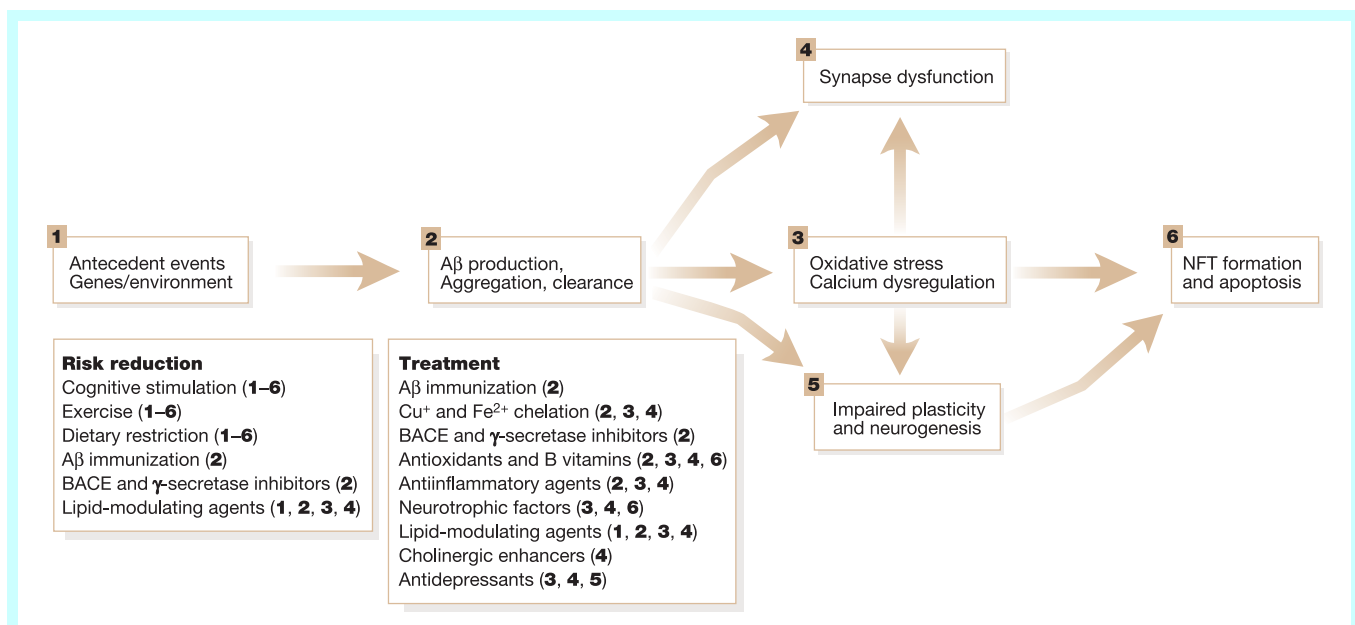


Figure 3 Strategies and targets for the prevention and treatment of AD. Approaches that are being tested in clinical and/or primary prevention trials include A β immunization, Cu⁺/Fe²⁺ chelation, cholesterol-lowering drugs (statins), anti-

inflammatory agents, antioxidants and folic acid. Epidemiological and animal studies suggest the potential benefits of cognitive stimulation, regular physical exercise and dietary restriction.

cholesterol metabolism might also promote neuronal degeneration by perturbing membrane fluidity and signal transduction. Increasing evidence suggests that diets high in saturated fats may increase the risk of AD, whereas diets rich in mono- and polyunsaturated fatty acids and omega-3 fatty acids may decrease disease risk, and several studies suggest that diets rich in omega-3 fatty acids, such as those found in fish, can reduce the risk of AD³⁵. Although the emerging data linking cholesterol and fatty acids to AD are encouraging, the potential of dietary modifications of fat intake to affect disease risk remains to be established.

Microdomains of plasma membranes called lipid rafts, in which the outer leaflet of the lipid bilayer is enriched in cholesterol and sphingomyelin may be sites at which several molecular events implicated in AD pathogenesis occur, including synaptic signal transduction, APP processing and the initiation of apoptosis³⁶. Analyses of raft lipids in brain tissue samples from AD and control subjects have demonstrated increased amounts of cholesterol and ceramides in AD, which are associated with increased levels of membrane-associated oxidative stress³⁷. Exposure of neurons to A β or other oxidative insults results in increased accumulation of cholesterol and ceramides in the neurons and these alterations can be prevented by treating the neurons with vitamin E and inhibitors of ceramide³⁷, suggesting that increased membrane oxidative stress may induce the production of cytotoxic ceramides in AD. Further studies in animal models and AD patients are required to clarify the roles of perturbed lipid metabolism in AD.

Synaptic and axonal alterations short circuit cognition

Learning and memory requires information exchange between neurons in circuits that pair at least two different environmental inputs in space and time. For example, when we meet someone new visual and auditory pathways are activated and this information is routed to the hippocampus, a brain region in which the convergent activity leads to enduring changes in synaptic strength so that the next time we see the person or hear his or her voice we remember his or her name. For reasons that remain elusive, neuronal circuits critical for learning and memory are particularly vulnerable to dysfunction and degeneration in AD. Emerging evidence implicates subtle changes in the function and structure of synapses and axons in these brain regions as being early and pivotal events in the pathogenesis of the neurodegenerative process in AD (Box 2). Consistent with this possibility, drugs that enhance activation of those synapses, such as acetylcholinesterase inhibitors, can improve cognition during the early stages of clinical disease³⁸.

A neuron's demise

Loss of neurons in the entorhinal cortex, hippocampus, frontal, parietal and temporal cortices of AD patients has been documented³⁹. Neurons in layer II of the entorhinal cortex and hippocampal CA1 neurons are particularly vulnerable. The reason(s) for this selective vulnerability remains uncertain, but might be related to the expression of genes that either promote or prevent neuronal death, including glutamate receptors, calcium-binding proteins and neurotrophic factors. The pattern of neuronal loss in AD overlaps with, but is not identical to, that of normal ageing—suggesting that AD pathogenesis is not simply an acceleration of normal brain ageing. Large numbers of neurofibrillary tangle-bearing neurons occur in AD, but neuronal loss exceeds the number of neurofibrillary tangles, which suggests that tangle-bearing neurons are removed and/or that some neurons die without forming tangles⁴⁰.

The death of populations of neurons in the brain regions affected in AD apparently occurs over a prolonged time period of many years, which suggests that a relatively small number of neurons are dying at any one time. Such a spatio-temporal pattern of cell death is characteristic of a form of programmed cell death called apoptosis and contrasts with necrosis in which cells die *en masse*. Increasing evidence suggests that many neurons may die by apoptosis in AD

(Box 3 and ref. 41), although it may not be the only form of cell death.

Of glia and microglia

Glial cells (astrocytes, oligodendrocytes and microglia) greatly outnumber neurons in the brain; they express many of the genes linked to AD and are subjected to many of the same environmental conditions as are neurons. Alterations in each of the three types of glial cells have been documented in studies of AD patients and of animal and cell culture models of AD, and available data suggest that A β plays a role in inducing many of the alterations in glial cells⁴². Abnormalities in astrocytes that may contribute to synaptic dysfunction and neuronal death include impaired glutamate transport, perturbed calcium regulation and production of pro-inflammatory cytokines. White matter consists of axons and oligodendrocytes, the cells that wrap around the axons, insulating them, and thereby facilitate rapid conduction of action potentials. Degeneration of cells in white matter is prominent in the brains of AD patients, and oligodendrocytes are vulnerable to being damaged and killed by A β ⁴³.

Microglia are macrophage-like cells that play important roles in responses of the brain to injury and infection. In AD, activated microglia congregate around amyloid plaques and degenerating neurons, and may produce toxins and inflammatory cytokines that contribute to the neurodegenerative process⁴². Both innate and cell-mediated immune mechanisms are involved in the pathogenesis of AD. In addition to activation of microglia, the innate immune response includes engagement of the classical complement cascade and induction of chemokines and pentraxins⁴². Evidence for participation of cell- and antibody-mediated responses in AD pathogenesis comes from recent studies demonstrating the ability of the immune system to generate antibodies against A β that may promote removal of A β from the brain^{44,45}.

Whereas many of the changes in glial cells in AD may promote neuronal degeneration, some of the changes may represent adaptive responses aimed at promoting neuronal plasticity and survival. For example, microglia may also remove A β , a potentially beneficial action of these immune cells⁴⁶. In addition, although synapses degenerate in vulnerable neuronal circuits, the remaining synapses may increase in size to compensate and astrocytes may play a role in this process^{47,48}. Moreover, the production of neurotrophic factors such as basic fibroblast growth may increase in astrocytes associated with A β deposits⁴⁹—such neurotrophic factors as well as certain cytokines⁵⁰ may counteract the neurodegenerative process in AD.

Pathways away from Alzheimer's disease

Diet and lifestyle

Potentially effective means of decreasing disease risk, and of inhibiting or even reversing the AD disease process in symptomatic patients, are emerging from recent studies (Fig. 3). Although not yet fully established, available data suggest that cognitively stimulating environments, physical exercise and diets low in calories and 'bad' fats (cholesterol and saturated fats) may reduce the risk of AD^{9,10,51}. The latter epidemiological findings are supported by animal studies showing that cognitively stimulating environments, physical exercise and dietary restriction regimens increase the resistance of neurons in the brain to degeneration, enhance neurogenesis and improve learning and memory^{10–12}. Exercise, cognitive stimulation and dietary restriction may each exert a beneficial effect through a similar mechanism involving increased production of brain-derived neurotrophic factor (BDNF)^{10–12}, although the importance of BDNF in AD remains to be determined. Vitamin supplementation may reduce disease risk, as suggested by a recent prospective study demonstrating reduced prevalence and incidence of AD in individuals taking vitamins C and E⁵², and the association of high folic acid levels and decreased homocysteine levels with reduced disease risk^{10,53}. The possibility that one's risk for AD can be reduced by

modifications of diet and lifestyle is of considerable interest, and suggests the potential for reducing the incidence of AD by preventative strategies similar to those that reduce the risk of cardiovascular disease.

Targeted therapeutics

Drugs are in development that target specific sites in the neurodegenerative cascade. Because the production of neurotoxic forms of A β from APP appears to be a pivotal event in AD pathogenesis, there is intense interest in developing drugs that block the β - or γ -secretase enzymes. Specific γ -secretase inhibitors that decrease A β production have been produced⁵⁴, but their use in humans may be compromised by side effects resulting from blockade of γ -secretase cleavage of Notch and other protein substrates. BACE inhibitors⁵⁵ may prove beneficial in reducing A β production without major side effects, because BACE-deficient mice have reduced A β production and do not exhibit any overt abnormal phenotypes⁵⁶. Another approach to reducing amyloid accumulation in the brain is agents that chelate copper and iron^{14,57}; such chelators would also be expected to reduce oxidative stress in neurons. The emerging link between cholesterol levels and AD has led to trials of cholesterol-lowering statins in AD patients (<http://www.alzforum.org/dis/tre/drc>).

One promising approach for preventing and treating AD is based upon stimulating the immune system to remove A β from the brain. The initial reports that immunization with human A β 42 (ref. 58) or passive immunization with A β antibodies⁵⁹ result in the clearance of A β plaques from the brains of APP mutant transgenic mice were followed by reports that such immunization approaches can also ameliorate memory deficits in the mice^{18,60}. The results of an initial clinical trial suggest that A β immunization may be an effective treatment for AD, although adverse reactions occurred in some patients and refinement of the immunization methods will be required⁶¹. In addition, the rate of cognitive decline in the AD group that received the vaccine but did not respond by plaque clearance was much greater than is typically seen in AD patients. Therefore, although the vaccine responders were cognitively better than the non-responders, their rate of cognitive decline was similar to that typical for patients with sporadic AD. It should also be noted that there have been many failures of trials in human patients of treatments that worked well in animal models of other neurodegenerative disorders, including Parkinson's disease and stroke. Nevertheless, an effective vaccine would have a major impact on the disease, and is currently one of the most exciting areas of AD research.

Other therapeutic approaches being tested include anti-inflammatory agents such as COX-2 inhibitors and aspirin⁶², and steroids that decline during normal ageing such as oestrogen and testosterone⁶³. Drugs that target specific sites in neurodegenerative cascades have, to date, only been employed in cell culture and animal models of AD, and their potential in the clinic remains unclear. Finally, the ability to identify individuals at risk for AD, based upon genetic (Apo E4 genotype, for example) or environmental (overweight sedentary lifestyle) factors, will allow the application of more aggressive interventions in those individuals. □

doi:10.1038/nature02621.

- Dickson, D. W. Neuropathological diagnosis of Alzheimer's disease: a perspective from longitudinal clinicopathological studies. *Neurobiol. Aging* **18**, S21–S26 (1997).
- Braak, H. & Braak, E. Evolution of neuronal changes in the course of Alzheimer's disease. *J. Neural Transm. Suppl.* **53**, 127–140 (1998).
- Selkoe, D. J. & Schenk, D. Alzheimer's disease: molecular understanding predicts amyloid-based therapeutics. *Annu. Rev. Pharmacol. Toxicol.* **43**, 545–584 (2003).
- Hardy, J. Amyloid, the presenilins and Alzheimer's disease. *Trends Neurosci.* **20**, 154–159 (1997).
- Sherrington, R. et al. Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. *Nature* **375**, 754–760 (1995).
- Levy-Lahad, E. et al. Candidate gene for the chromosome 1 familial Alzheimer's disease locus. *Science* **269**, 973–977 (1995).
- Roses, A. D. A model for susceptibility polymorphisms for complex diseases: apolipoprotein E and Alzheimer disease. *Neurogenetics* **1**, 3–11 (1997).
- Tanzi, R. E. & Bertram, L. New frontiers in Alzheimer's disease genetics. *Neuron* **32**, 181–184 (2001).
- Mayeux, R. Epidemiology of neurodegeneration. *Annu. Rev. Neurosci.* **26**, 81–104 (2003).
- Mattson, M. P. Gene-diet interactions in brain aging and neurodegenerative disorders. *Ann. Intern. Med.* **139**, 441–444 (2003).
- Young, D. et al. Environmental enrichment inhibits spontaneous apoptosis, prevents seizures and is neuroprotective. *Nature Med.* **5**, 448–453 (1999).
- Lee, J., Duan, W. & Mattson, M. P. Evidence that brain-derived neurotrophic factor is required for basal neurogenesis and mediates, in part, the enhancement of neurogenesis by dietary restriction in the hippocampus of adult mice. *J. Neurochem.* **82**, 1367–1375 (2002).
- Cotman, C. W. & Berchtold, N. C. Exercise: a behavioral intervention to enhance brain health and plasticity. *Trends Neurosci.* **25**, 295–301 (2002).
- Bush, A. I., Masters, C. L. & Tanzi, R. E. Copper, beta-amyloid, and Alzheimer's disease: tapping a sensitive connection. *Proc. Natl Acad. Sci. USA* **100**, 11193–11194 (2003).
- Haass, C. et al. The Swedish mutation causes early-onset Alzheimer's disease by beta-secretase cleavage within the secretory pathway. *Nature Med.* **1**, 1291–1296 (1995).
- Scheuner, D. et al. Secreted amyloid beta-protein similar to that in senile plaques of Alzheimer's disease is increased *in vivo* by the presenilin 1 and 2 APP mutations linked to familial Alzheimer's disease. *Nature Med.* **2**, 864–870 (1996).
- Mattson, M. P. Cellular actions of beta-amyloid precursor protein and its soluble and fibrillogenic derivatives. *Physiol. Rev.* **77**, 1081–1132 (1997).
- Morgan, D. et al. A beta peptide vaccination prevents memory loss in an animal model of Alzheimer's disease. *Nature* **408**, 982–985 (2000).
- Butterfield, D. A. et al. Evidence of oxidative damage in Alzheimer's disease brain: central role for amyloid beta-peptide. *Trends Mol. Med.* **7**, 548–554 (2001).
- Smith, M. A. et al. Iron accumulation in Alzheimer disease is a source of redox-generated free radicals. *Proc. Natl Acad. Sci. USA* **94**, 9866–9868 (1997).
- Blass, J. P. Brain metabolism and brain disease: is metabolic deficiency the proximate cause of Alzheimer dementia? *J. Neurosci. Res.* **66**, 851–856 (2001).
- Dodart, J. C. et al. Early regional cerebral glucose hypometabolism in transgenic mice overexpressing the V717F beta-amyloid precursor protein. *Neurosci. Lett.* **277**, 49–52 (1999).
- Buchner, M., Huber, R., Sturchler-Pierrat, C., Staufenbiel, M. & Riepe, M. W. Impaired hypoxic tolerance and altered protein binding of NADH in presymptomatic APP23 transgenic mice. *Neuroscience* **114**, 285–289 (2002).
- Watson, G. S. & Craft, S. The role of insulin resistance in the pathogenesis of Alzheimer's disease: implications for treatment. *CNS Drugs* **17**, 24–35 (2003).
- Gabuzda, D., Busciglio, J., Chen, L. B., Matsudaira, P. & Yankner, B. A. Inhibition of energy metabolism alters the processing of amyloid precursor protein and induces a potentially amyloidogenic derivative. *J. Biol. Chem.* **269**, 13623–13628 (1994).
- Saito, K. et al. Widespread activation of calcium-activated neutral proteinase (calpain) in the brain in Alzheimer disease: a potential molecular basis for neuronal degeneration. *Proc. Natl Acad. Sci. USA* **90**, 2628–2632 (1993).
- Mattson, M. P. & Chan, S. L. Neuronal and glial calcium signaling in Alzheimer's disease. *Cell Calcium* **34**, 385–397 (2003).
- Le, Y. et al. Amyloid (beta)42 activates a G-protein-coupled chemoattractant receptor, FPR-like-1. *J. Neurosci.* **21**, RC123 (2001).
- LaFerla, F. M. Calcium dyshomeostasis and intracellular signaling in Alzheimer's disease. *Nature Rev. Neurosci.* **3**, 862–872 (2002).
- Yang, Y. & Cook, D. G. Presenilin-1 deficiency impairs glutamate-evoked intracellular calcium responses in neurons. *Neuroscience* **124**, 501–506 (2004).
- Eckert, A. et al. Alzheimer's disease-like alterations in peripheral cells from presenilin-1 transgenic mice. *Neurobiol. Dis.* **8**, 331–342 (2001).
- Pugliese, L., Tanzi, R. E. & Kovacs, D. M. Alzheimer's disease: the cholesterol connection. *Nature Neurosci.* **6**, 345–351 (2003).
- Wolozin, B. et al. Decreased prevalence of Alzheimer disease associated with 3-hydroxy-3-methylglutaryl coenzyme A reductase inhibitors. *Arch. Neurol.* **57**, 1439–1443 (2000).
- Fassbender, K. et al. Simvastatin strongly reduces levels of Alzheimer's disease beta-amyloid peptides Abeta 42 and Abeta 40 *in vitro* and *in vivo*. *Proc. Natl Acad. Sci. USA* **98**, 5856–5861 (2001).
- Kalmijn, S. Fatty acid intake and the risk of dementia and cognitive decline: a review of clinical and epidemiological studies. *J. Nutr. Health Aging* **4**, 202–207 (2000).
- Ehehalt, R. et al. Amyloidogenic processing of the Alzheimer beta-amyloid precursor protein depends on lipid rafts. *J. Cell Biol.* **160**, 113–123 (2003).
- Cutler, R. G. et al. Involvement of perturbed ceramide and cholesterol metabolism in brain aging and Alzheimer's disease. *Proc. Natl Acad. Sci. USA* **101**, 2070–2075 (2004).
- Farlow, M. A clinical overview of cholinesterase inhibitors in Alzheimer's disease. *Int. Psychogeriatr.* **14**, 93–126 (2002).
- West, M. J. et al. Differences in the pattern of hippocampal neuronal loss in normal ageing and Alzheimer's disease. *Lancet* **344**, 769–772 (1994).
- Gomez-Isla, T. et al. Neuronal loss correlates with but exceeds neurofibrillary tangles in Alzheimer's disease. *Ann. Neurol.* **41**, 17–24 (1997).
- Lustbader, J. W. et al. A β AD directly links A β to mitochondrial toxicity in Alzheimer's disease. *Science* **304**, 448–452 (2004).
- McGeer, P. L. & McGeer, E. G. Local neuroinflammation and the progression of Alzheimer's disease. *J. Neurovirol.* **8**, 529–538 (2002).
- Xu, J. et al. Amyloid-beta peptides are cytotoxic to oligodendrocytes. *J. Neurosci.* **21**, RC118 (2001).
- Schenk, D. et al. Immunization with amyloid-beta attenuates Alzheimer-disease-like pathology in the PDAPP mouse. *Nature* **400**, 173–177 (1999).
- Bard, F. et al. Peripherally administered antibodies against amyloid beta-peptide enter the central nervous system and reduce pathology in a mouse model of Alzheimer disease. *Nature Med.* **6**, 916–919 (2000).
- Jantzen, P. T. et al. Microglial activation and beta-amyloid deposit reduction caused by a nitric oxide-releasing nonsteroidal anti-inflammatory drug in amyloid precursor protein plus presenilin-1 transgenic mice. *J. Neurosci.* **22**, 2246–2254 (2002).
- DeKosky, S. T. & Scheff, S. W. Synapse loss in frontal cortex biopsies in Alzheimer's disease: correlation with cognitive severity. *Ann. Neurol.* **27**, 457–464 (1990).

48. Murai, K. K. *et al.* Control of hippocampal dendritic spine morphology through ephrin-A3/EphA4 signaling. *Nature Neurosci.* **6**, 153–160 (2003).
49. Cummings, B. J. *et al.* Neuritic involvement within bFGF immunopositive plaques of Alzheimer's disease. *Exp. Neurol.* **124**, 315–325 (1993).
50. Barger, S. W. *et al.* Tumor necrosis factors alpha and beta protect neurons against amyloid beta-peptide toxicity: evidence for involvement of a kappa B-binding factor and attenuation of peroxide and Ca²⁺ accumulation. *Proc. Natl Acad. Sci. USA* **92**, 9328–9332 (1995).
51. Pope, S. K., Shue, V. M. & Beck, C. Will a healthy lifestyle help prevent Alzheimer's disease? *Annu. Rev. Public Health* **24**, 111–132 (2003).
52. Zandi, P. P. *et al.* Reduced risk of Alzheimer disease in users of antioxidant vitamin supplements: the Cache County Study. *Arch. Neurol.* **61**, 82–88 (2004).
53. Wang, H. X. *et al.* Vitamin B(12) and folate in relation to the development of Alzheimer's disease. *Neurology* **56**, 1188–1194 (2001).
54. Dewachter, L. & Van Leuven, F. Secretases as targets for the treatment of Alzheimer's disease: the prospects. *Lancet Neurol.* **1**, 409–416 (2002).
55. John, V. *et al.* Human beta-secretase (BACE) and BACE inhibitors. *J. Med. Chem.* **46**, 4625–4630 (2003).
56. Roberds, S. L. *et al.* BACE knockout mice are healthy despite lacking the primary beta-secretase activity in brain: implications for Alzheimer's disease therapeutics. *Hum. Mol. Genet.* **10**, 1317–1324 (2001).
57. Ritchie, C. W. *et al.* Metal-protein attenuation with iodochlorhydroxyquin (clioquinol) targeting Abeta amyloid deposition and toxicity in Alzheimer disease: a pilot phase 2 clinical trial. *Arch. Neurol.* **60**, 1685–1691 (2003).
58. Schenk, D. *et al.* Immunization with amyloid-beta attenuates Alzheimer-disease-like pathology in the PDAPP mouse. *Nature* **400**, 173–177 (1999).
59. McLaurin, J. *et al.* Therapeutically effective antibodies against amyloid-beta peptide target amyloid-beta residues 4–10 and inhibit cytotoxicity and fibrillogenesis. *Nature Med.* **8**, 1263–1269 (2002).
60. Kotilinek, L. A. *et al.* Reversible memory loss in a mouse transgenic model of Alzheimer's disease. *J. Neurosci.* **22**, 6331–6335 (2002).
61. Hock, C. *et al.* Antibodies against beta-amyloid slow cognitive decline in Alzheimer's disease. *Neuron* **38**, 547–554 (2003).
62. Hoozemans, J. J. *et al.* Non-steroidal anti-inflammatory drugs and cyclooxygenase in Alzheimer's disease. *Curr. Drug Targets* **4**, 461–468 (2003).
63. Resnick, S. M. & Henderson, V. W. Hormone therapy and risk of Alzheimer disease. *J. Am. Med. Assoc.* **288**, 2170–2172 (2002).
64. Buxbaum, J. D. *et al.* Evidence that tumor necrosis factor alpha converting enzyme is involved in regulated alpha-secretase cleavage of the Alzheimer amyloid protein precursor. *J. Biol. Chem.* **273**, 27765–27767 (1998).
65. Asai, M. *et al.* Putative function of ADAM9, ADAM10, and ADAM17 as APP alpha-secretase. *Biochem. Biophys. Res. Commun.* **301**, 231–235 (2003).
66. Leissring, M. A. *et al.* A physiologic signaling role for the gamma-secretase-derived intracellular fragment of APP. *Proc. Natl Acad. Sci. USA* **99**, 4697–4702 (2002).
67. Lu, D. C. *et al.* A second cytotoxic proteolytic peptide derived from amyloid beta-protein precursor. *Nature Med.* **6**, 397–404 (2000).
68. Yu, G. *et al.* Nicastrin modulates presenilin-mediated notch/glp-1 signal transduction and betaAPP processing. *Nature* **407**, 48–54 (2000).
69. Francis, R. *et al.* aph-1 and pen-2 are required for Notch pathway signaling, gamma-secretase cleavage of betaAPP, and presenilin protein accumulation. *Dev. Cell* **3**, 85–97 (2002).
70. Takasugi, N. *et al.* The role of presenilin cofactors in the gamma-secretase complex. *Nature* **422**, 438–441 (2003).
71. Shen, J. *et al.* Skeletal and CNS defects in Presenilin-1-deficient mice. *Cell* **89**, 629–639 (1997).
72. Lazarov, O., Lee, M., Peterson, D. A. & Sisodia, S. S. Evidence that synaptically released beta-amyloid accumulates as extracellular deposits in the hippocampus of transgenic mice. *J. Neurosci.* **22**, 9785–9793 (2002).
73. Kimberly, W. T. *et al.* The intracellular domain of the beta-amyloid precursor protein is stabilized by Fe65 and translocates to the nucleus in a notch-like manner. *J. Biol. Chem.* **276**, 40288–40292 (2001).
74. Yankner, B. A. *et al.* Neurotrophic and neurotoxic effects of amyloid beta protein: reversal by tachykinin neuropeptides. *Science* **250**, 279–282 (1990).
75. Lambert, M. P. *et al.* Diffusible, nonfibrillar ligands derived from Abeta1–42 are potent central nervous system neurotoxins. *Proc. Natl Acad. Sci. USA* **95**, 6448–6453 (1998).
76. Kaye, R. *et al.* Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. *Science* **300**, 486–489 (2003).
77. Chapman, P. F. *et al.* Impaired synaptic plasticity and learning in aged amyloid precursor protein transgenic mice. *Nature Neurosci.* **2**, 271–276 (1999).
78. Koistinaho, M. *et al.* Specific spatial learning deficits become severe with age in beta-amyloid precursor protein transgenic mice that harbor diffuse beta-amyloid deposits but do not form plaques. *Proc. Natl Acad. Sci. USA* **98**, 14675–14680 (2001).
79. Kamenetz, F. *et al.* APP processing and synaptic function. *Neuron* **37**, 925–937 (2003).
80. Gong, Y. *et al.* Alzheimer's disease-affected brain: presence of oligomeric A beta ligands (ADDLs) suggests a molecular basis for reversible memory loss. *Proc. Natl Acad. Sci. USA* **100**, 10417–10422 (2003).
81. Klein, W. L., Krafft, G. A. & Finch, C. E. Targeting small Aβ oligomers: the solution to an Alzheimer's disease conundrum? *Trends Neurosci.* **24**, 219–224 (2001).
82. Oster-Granite, M. L. *et al.* Age-dependent neuronal and synaptic degeneration in mice transgenic for the C terminus of the amyloid precursor protein. *J. Neurosci.* **16**, 6732–6741 (1996).
83. Oddo, S. *et al.* Triple-transgenic model of Alzheimer's disease with plaques and tangles: intracellular Aβeta and synaptic dysfunction. *Neuron* **39**, 409–421 (2003).
84. Mucke, L. *et al.* High-level neuronal expression of abeta 1–42 in wild-type human amyloid protein precursor transgenic mice: synaptotoxicity without plaque formation. *J. Neurosci.* **20**, 4050–4058 (2000).
85. Chan, S. L., Furukawa, K. & Mattson, M. P. Presenilins APP in neuritic and synaptic plasticity: implications for the pathogenesis of Alzheimer's disease. *Neuromolecular Med.* **2**, 167–196 (2002).
86. Stahl, R. *et al.* Assessment of axonal degeneration on Alzheimer's disease with diffusion tensor MRI. *Radiology* **43**, 566–575 (2003).
87. Buxbaum, J. D. *et al.* Alzheimer amyloid protein precursor in the rat hippocampus: transport and processing through the perforant path. *J. Neurosci.* **18**, 9629–9637 (1998).
88. Mandelkow, E. M. *et al.* Clogging of axons by tau, inhibition of axonal traffic and starvation of synapses. *Neurobiol. Aging* **24**, 1079–1085 (2003).
89. Lee, V. M., Goedert, M. & Trojanowski, J. Q. Neurodegenerative tauopathies. *Annu. Rev. Neurosci.* **24**, 1121–1159 (2001).
90. Masliah, E. *et al.* Altered expression of synaptic proteins occurs early during progression of Alzheimer's disease. *Neurology* **56**, 127–129 (2001).
91. Yao, P. J. Synaptic frailty and clathrin-mediated synaptic vesicle trafficking in Alzheimer's disease. *Trends Neurosci.* **27**, 24–29 (2004).
92. Yuan, J. & Yankner, B. A. Apoptosis in the nervous system. *Nature* **407**, 802–809 (2000).
93. Mattson, M. P. Apoptosis in neurodegenerative disorders. *Nature Rev. Mol. Cell Biol.* **1**, 120–129 (2000).
94. Eckert, A. *et al.* Mitochondrial dysfunction, apoptotic cell death, and Alzheimer's disease. *Biochem. Pharmacol.* **66**, 1627–1634 (2003).
95. McPhie, D. L. *et al.* DNA synthesis and neuronal apoptosis caused by familial Alzheimer disease mutants of the amyloid precursor protein are mediated by the p21 activated kinase PAK3. *J. Neurosci.* **23**, 6914–6927 (2003).
96. Lee, M. S. *et al.* Neurotoxicity induces cleavage of p35 to p25 by calpain. *Nature* **405**, 360–364 (2000).
97. Kruman, I. I. *et al.* Cell cycle activation linked to neuronal cell death initiated by DNA damage. *Neuron* **41**, 549–561 (2004).
98. Van Leeuwen, F. W. *et al.* Frameshift mutants of beta amyloid precursor protein and ubiquitin-B in Alzheimer's and Down patients. *Science* **279**, 242–247 (1998).
99. Song, S. *et al.* Essential role of E2–25K/Hip-2 in mediating amyloid-beta neurotoxicity. *Mol. Cell* **12**, 553–563 (2003).
100. Pedersen, W. A., Wan, R., Zhang, P. & Mattson, M. P. Urocortin, but not urocortin II, protects cultured hippocampal neurons from oxidative and excitotoxic cell death via corticotropin-releasing hormone receptor type I. *J. Neurosci.* **22**, 404–412 (2002).

Competing interests statement The author declares that he has no competing financial interests.

Correspondence and requests for materials should be addressed to M.P.M. (mattsonm@grc.nia.nih.gov).

Mechanism of transfer RNA maturation by CCA-adding enzyme without using an oligonucleotide template

Yong Xiong^{1,3} & Thomas A. Steitz^{1,2,3}

¹Department of Molecular Biophysics and Biochemistry, ²Department of Chemistry, Yale University and ³Howard Hughes Medical Institute, New Haven, Connecticut 06520, USA

Transfer RNA nucleotidyltransferases (CCA-adding enzymes) are responsible for the maturation or repair of the functional 3' end of tRNAs by means of the addition of the essential nucleotides CCA. However, it is unclear how tRNA nucleotidyltransferases polymerize CCA onto the 3' terminus of immature tRNAs without using a nucleic acid template. Here we describe the crystal structure of the *Archaeoglobus fulgidus* tRNA nucleotidyltransferase in complex with tRNA. We also present ternary complexes of this enzyme with both RNA duplex mimics of the tRNA acceptor stem that terminate with the nucleotides C74 or C75, as well as the appropriate incoming nucleoside 5'-triphosphates. A single nucleotide-binding pocket exists whose specificity for both CTP and ATP is determined by the protein side chain of Arg 224 and backbone phosphates of the tRNA, which are non-complementary to and thus exclude UTP and GTP. Discrimination between CTP or ATP at a given addition step and at termination arises from changes in the size and shape of the nucleotide binding site that is progressively altered by the elongating 3' end of the tRNA.

Transfer RNA nucleotidyltransferases (CCA-adding enzymes) catalyse the post-transcriptional addition of CCA onto the 3' terminus of immature tRNAs using CTP and ATP as substrates¹. Although the CCA end of tRNA is universally conserved and absolutely required for protein biosynthesis^{2,3}, it is not encoded in many eubacterial or archaeal tRNA genes, or in nearly all eukaryotic tRNA genes⁴. In the few organisms that do have the CCA sequence included in their genes for tRNA (such as *Escherichia coli*), the CCA-adding enzyme serves to maintain the integrity of the CCA terminus of tRNA⁵.

CCA-adding enzymes belong to the nucleotidyltransferase superfamily⁶ whose members are divided into two classes on the basis of their sequences⁷. Examples of class I enzymes include the archaeal CCA-adding enzyme, DNA polymerase β (Pol β), eukaryotic poly(A) polymerase (PAP) and terminal deoxynucleotidyltransferase (TdT). Class II enzymes include the eubacterial and eukaryotic CCA-adding enzymes and eubacterial PAP. Crystal structures of CCA-adding enzymes from both classes confirm that they share only a common amino-terminal catalytic domain, which is also homologous to the polymerase domain found in all nucleotidyltransferase family enzymes^{8–11}.

The ability of the CCA-adding enzyme to add specific nucleotides in the absence of a nucleic acid template makes it one of the most intriguing polymerases. Currently, only PAP, TdT and CCA-adding enzymes are known to function without a nucleic acid template, and between them the CCA-adding enzymes exhibit the highest degree of specificity for both sequence and length. Biochemical studies, sequence analyses and structural data have established that the underlining chemistry for nucleotide addition is the same two-metal-ion-catalysed reaction used by all polymerases¹²; however, insights into the source of the CCA-adding enzyme's specificity of nucleotide incorporation are scarce.

Several models explaining how the enzyme incorporates CTP and ATP with high specificity have been proposed. Early models proposed multiple binding sites for CTP and ATP^{13–15}, but these were excluded owing to the existence of a single nucleotidyltransferase active site in the sequence^{6,16} and observations of a single nucleotide-binding site in the crystal structures of both class I and class II CCA-adding enzymes^{8,10}. Subsequent models were based on additional biochemical and structural data. Both the 'collaborative

templating model' and the 'scrunching model' postulated a progressive refolding of the growing CCA terminus of the tRNA, which allows for the re-use of the single nucleotide-binding pocket while altering the templating specificity of the enzyme^{8,17}. The structure of the class II CCA-adding enzyme of *Bacillus stearothermophilus* (BstCCA) binds both CTP and ATP, but excludes UTP and GTP, by binding to the Watson–Crick base edge of both nucleotides using the same protein side chains⁸. Recently, a model of the complex between a class I CCA-adding enzyme and tRNA was proposed¹⁰ that required the 3' terminus of tRNA to unstack from the acceptor stem in order to position the priming nucleotide correctly for reaction with the bound NTP. In contrast to the conclusion of another study¹¹, this model also required the binding of the tRNA substrate for nucleotide selection, as all NTPs bound to the enzyme in its absence¹⁰.

Although these latter models could explain certain aspects of the experimental observations, the central question regarding the specificity of the CCA-adding enzyme remained unanswered: how does the CCA-adding enzyme achieve its specificity of nucleotide incorporation without using a nucleic acid template? The crystal structures of the class I CCA-adding enzyme of *A. fulgidus* (AfCCA) in complex with tRNA, as well as its complexes with two RNA duplex mimics of the tRNA acceptor stem that end with the nucleotides C74 or C75 in the presence of the appropriate incoming nucleotide triphosphates, now provide the structural basis of this specificity.

Structure determination

The co-crystal structure of AfCCA bound to full-length yeast tRNA^{Phe} was determined at 6.5 Å resolution by the molecular replacement method using the structure of the dimeric apo-AfCCA as a search model¹⁰. As a consequence of the unusually high solvent content of ~84% in the crystal, cycles of solvent electron density flattening were extremely effective in improving the electron density map and, together with two-fold non-crystallographic symmetry (NCS) averaging, produced a map of exceptional quality. All parts of the complex including the carboxy-terminal histidine tag of the AfCCA enzyme, which was not included in the model, exhibited well-defined electron density

(see Supplementary Fig. S1). The 3' terminus of the tRNA was built into an unbiased electron density map that was calculated before its inclusion in the model (Supplementary Fig. 2a). The structure of this complex was refined as groups of rigid bodies to a free *R*-factor of 41%.

AfCCA was also crystallized with synthetic RNA duplexes mimicking the acceptor stem of the tRNA: they have the nucleotide corresponding to either C74 or C75 at their 3' termini, and are named AC74 and ACC75, respectively (Fig. 1c). Self-complementary RNA oligonucleotides were used to ensure the binding of the appropriate terminus to the active site as well as the formation of homogeneous duplexes in solution. The ternary complexes with incoming nucleotides were prepared by soaking ATP or CTP into crystals of the binary complexes, or by co-crystallizing the enzyme with ACC75 and a non-reactive $\alpha\beta$ -methylene ATP. The crystal structures of the AC74 and ACC75 ternary complexes were determined by molecular replacement using the *apo*-AfCCA dimer as a search model. The initial phases derived from the protein model were improved by NCS averaging within each crystal and by cross-crystal averaging between highly non-isomorphous ternary complex crystals as well as the *apo*-AfCCA crystal¹⁰. The structures of the incoming nucleotides and the 3' terminal nucleotides were built into unbiased electron density maps that were calculated before their inclusions in the models (Supplementary Fig. S2b, c). The final models were refined to free *R*-factors of 29.1% and 24.1% at resolutions of 3.4 Å and 2.2 Å, respectively (Table 1).

Overall structures of the complexes

Two tRNA molecules bind to the AfCCA dimer with each tRNA contacting only one subunit (Fig. 1a, b). As predicted in our earlier model¹⁰, the tRNA acceptor stem and T-stem loop are inserted into the extended cleft on the enzyme, with the CCA terminus being in the active site of the head domain, the T loop contacting the tail domain and the anticodon stem loop pointing away from the enzyme (Supplementary Movie 1). Here, the head, neck, body and tail domains of the enzyme are defined as in the *apo*-AfCCA structure¹⁰. The role of the body and tail domains of the class I enzyme in binding tRNA supports the proposal¹⁰ that the corresponding domains also bind to tRNA in class II enzymes because their shape and dimensions are similar, and they are complementary to the shape of the acceptor and T stems of tRNA. The duplex RNA constructs that mimic the acceptor stem of the tRNA bind to AfCCA in an almost identical fashion to tRNA, and they superimpose on the bound tRNA particularly well in the regions that make contact with the enzyme (Fig. 1d, e), consistent with the observation that a tRNA mini-helix functions as an efficient substrate for CCA addition¹⁸. The enzyme does not add CCA to other RNAs in the cell, presumably because they do not form a simple duplex structure of appropriate length containing a suitable 3' nucleotide overhang.

The acceptor stems of the AC74 and ACC75 mimics, as well as the tRNA product, bind in the same location and orientation on the enzyme, in agreement with the conclusion derived from biochemical crosslinking studies that the tRNA does not translocate during

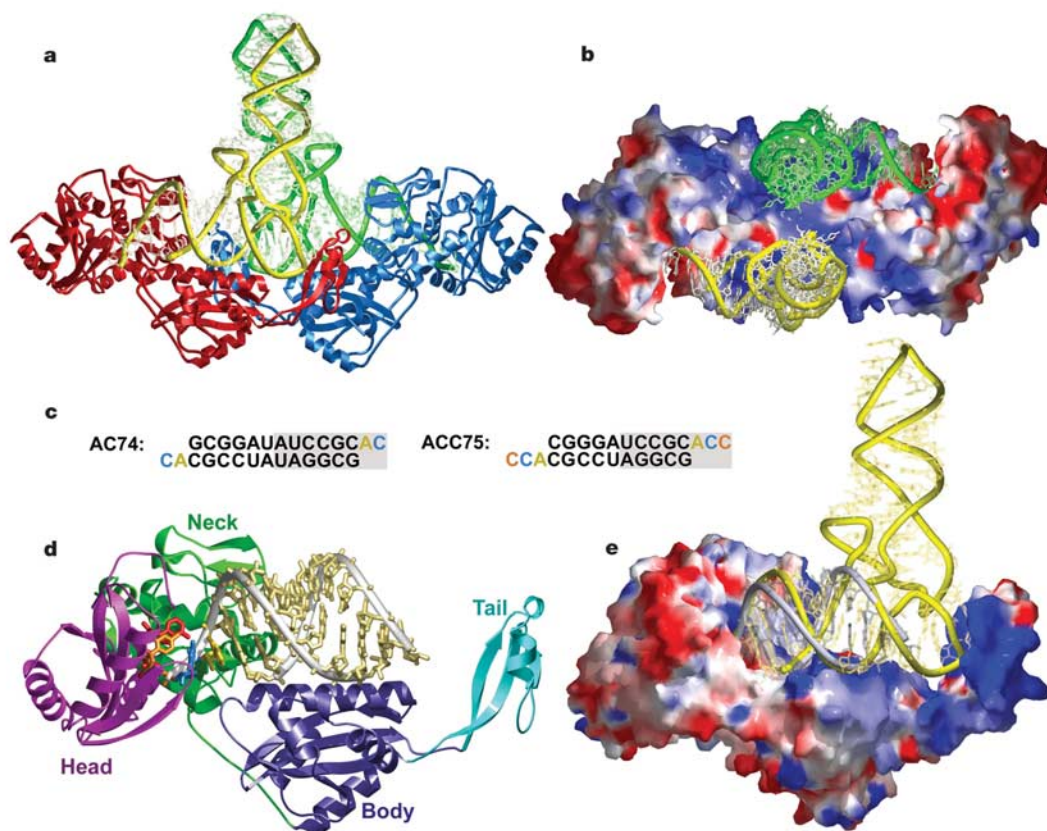


Figure 1 Overall structures of the AfCCA substrate complexes. **a**, The AfCCA–tRNA dimer. The tRNA backbone is highlighted with yellow and green coils, whereas AfCCA subunits are represented by red and blue ribbons. **b**, View of the AfCCA–tRNA dimer rotated by 90° about a horizontal axis from that in **a**. AfCCA is shown in a surface representation, with blue for positive and red for negative electrostatic potentials. **c**, The sequences of the AC74 and the ACC75 RNA constructs. The boxed regions indicate

sequences identical to those of the acceptor stem of yeast tRNA^{phe}. **d**, The structure of AfCCA in complex with ACC75. The RNA is in a stick representation whereas the enzyme is in a ribbon representation with head, neck, body and tail domains coloured magenta, green, blue and cyan, respectively. **e**, Superposition of the complexes with tRNA in yellow and the ACC75 duplex in white.

CCA addition¹⁷. Comparison of the protein structures of these three complexes with that of the *apo*-AfCCA shows that there are no large conformational changes in the enzyme except for a small repositioning of the N-terminal head domain. Except for the 3'-terminal CCA bases, the enzyme interacts almost exclusively with the sugar-phosphate backbone of the tRNA through electrostatic, hydrogen-bonding and hydrophobic interactions, which is consistent with the enzyme's ability to bind any tRNA molecule in a sequence-independent manner. The only hydrogen bond to a base, which is between the O2 atom of C72 and Tyr 169, is nonspecific.

Snapshots of the CCA addition process

These three crystal structures represent snapshots of three steps that occur during CCA addition (Supplementary Movie 2). Throughout this process, the acceptor stem along with the discriminator base A73 remains fixed on the enzyme, whereas the growing 3' terminus of the tRNA refolds to position the primer terminal nucleotide identically at each step, and allow the identical binding of the incoming CTP or ATP in a single pocket (Fig. 2). As anticipated in our earlier modelling study¹⁰, the base of the 3'-terminal nucleotide is not stacked on the acceptor stem bases. Rather, for both AC74 and ACC75, the base of the primer nucleotide flips about 90° from its stacking position on the acceptor stem and stacks onto the base of the incoming nucleoside triphosphate.

The structure of AfCCA in complex with AC74 and incoming CTP shows the base of the primer nucleotide C74 stacking on the base of the incoming CTP with its 3' hydroxyl groups positioned to attack the α -phosphate of CTP (Fig. 2a). C74 is disordered without the incoming CTP in the crystal structure of the binary AfCCA-AC74 complex (data not shown). At this step in the process, the flat plane of the peptide connecting Ala 95 and Glu 96 and lying at the apex of a β -turn region in the AfCCA head domain stacks onto the base of A73.

Table 1 Data and refinement statistics

Complex	AC74	ACC75	tRNA ^{pho}
Space group	<i>P</i> 2 ₁	<i>P</i> 2 ₁	<i>P</i> 321
Resolution (Å)	3.4 (3.6–3.4)	2.2 (2.3–2.2)	6.5 (6.7–6.5)
<i>R</i> _{sym} (%)	9.8 (58.5)	7.2 (73.4)	10.2
Completeness (%)	99.0 (99.4)	98.8 (97.5)	99.2 (97.6)
<i>I</i> / σ	10.2 (1.6)	15.2 (1.5)	12.2 (1.5)
<i>R</i> _{work} / <i>R</i> _{free} (%)	22.7/29.1	21.7/24.1	41.0/41.3
r.m.s.d. (bond/angle)	0.016/1.8	0.017/1.7	—

Cell constants for the *P*2₁ crystals are: *a* = 116.6 Å, *b* = 84.8 Å, *c* = 135.8 Å, β = 104.7°. Cell constants for the *P*321 crystal are: *a* = *b* = 218.4 Å, *c* = 170.3 Å, α = β = 90°, γ = 120°. r.m.s.d., root-mean-square deviation. Values in parentheses are statistics for the highest resolution shell.

After C75 has been incorporated, C74 moves from the primer position to a location closer to the acceptor stem (Fig. 2b); however, instead of stacking on the acceptor stem, the base of C74 forms four hydrogen bonds between its Watson–Crick base edge and its 2'-OH group and protein residues 95–99 located in the β -turn. The newly incorporated C75 shifts into the primer position, with its 3' hydroxyl group pointed towards the α -phosphate of the newly arrived ATP, ready for the next round of nucleotide addition.

Upon the incorporation of A76, the CCA terminus adopts a continuously stacked helical conformation emanating from the acceptor stem, with the base surface of A76 packed against the β -sheet in the head domain of AfCCA (Fig. 2c).

Active site geometry and the nucleotide-binding pocket

The conformations of the catalytic site and the bound substrates observed in the current crystal structures (Fig. 3a) are nearly identical to the corresponding structure of the ternary complex of DNA Pol β ¹⁹, which is also a class I nucleotidyltransferase. When the structure of the AC74 complex is aligned on that of the Pol β ternary complex by superimposing the catalytic carboxylates as well as the triphosphates of the incoming nucleotides, the primer terminal nucleotides of both complexes are also superimposed. Although the duplexes containing the primer nucleotide are bound to these two enzymes with orientations that differ by almost 90° (Fig. 3b), the primer termini are presented to the catalytic site in the same way. The class I AfCCA achieves this by unstacking the 3'-terminal nucleotide from the acceptor stem of the tRNA and rotating it by nearly 90°. In class II BstCCA, the modelled tRNA acceptor stem¹⁰ approaches the homologous catalytic domain from a direction that is perpendicular to both the duplex substrates bound to Pol β and AfCCA (Fig. 3b). As a consequence, the primer nucleotide must also unstack and flip about 90° to be positioned in the primer position observed in both the Pol β and AfCCA ternary complexes.

Two divalent metal ions (A and B) are observed bound to the catalytic carboxylates and the 3' terminus of the primer strand of AC74 (Fig. 3a), in agreement with the proposal that all polymerases use a common two-metal-ion mechanism for their nucleotide incorporation¹². Only one metal ion (B) is observed in the ACC75 complex, which has its 3' terminus in a slightly different conformation (Supplementary Fig. 3). This is presumably due to the experimental conditions (see Methods).

The nucleotide-binding pocket for the incoming CTP or ATP is formed by residues from the head and neck domains of the enzyme as well as by backbone phosphates located near the 3' terminus of the tRNA (Fig. 2). This observation that tRNA forms a component of the nucleotide-binding site explains why *apo*-AfCCA is unable to distinguish between correct and incorrect NTPs and why the base is

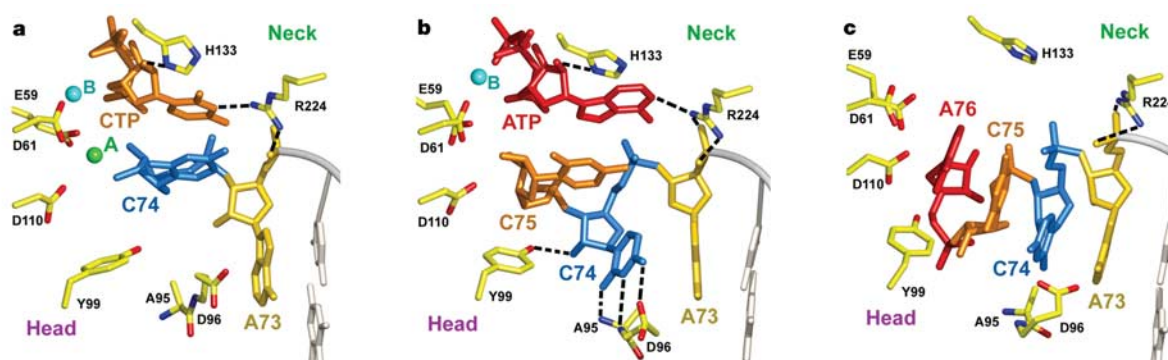


Figure 2 The progressive refolding of the 3' terminus. The nucleotides A73, C74, C75 and A76 are coloured yellow, blue, orange and red, respectively. Selected protein side chains are shown with hydrogen-bonding interactions represented by dashed lines. **a**, The AC74 complex with the incoming CTP (orange). The cyan sphere represents metal

ion B and the green sphere represents metal ion A, which is observed in a different Mn²⁺ ion-soaked crystal. **b**, The ACC75 complex with the incoming ATP (red). **c**, The tRNA complex.

not ordered in the binary complex¹⁰. The triphosphate moieties of both CTP and ATP interact extensively with the enzyme, as is the case in the binary complexes without tRNA. The 2'-OH group also forms a hydrogen bond to His 133 in both complexes.

The bases of the incoming nucleotides are recognized by both the enzyme and the tRNA by means of hydrogen-bonding interactions

to their Watson–Crick edges. The exocyclic N4 atom of CTP forms two hydrogen bonds with the tRNA backbone phosphate of A73 and C74, whereas the N3 atom of CTP is recognized by the Arg 224 guanidinium group of AfCCA (Fig. 4a). Similarly, the N6 of ATP forms hydrogen bonds with phosphate oxygens of A73 and C74, and the N1 atom also interacts with Arg 224 (Fig. 4b). This arginine,

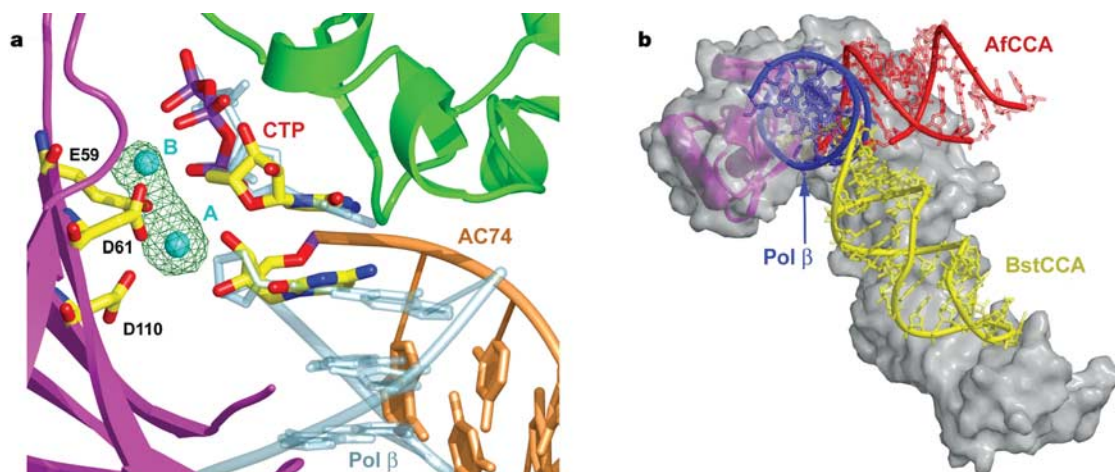


Figure 3 The two-metal-ion geometry and orientations of primer duplexes in the two classes of CCA-adding enzyme and in Pol β. **a**, Superimposition of the active site of the AC74 complex and that of Pol β. AfCCA is colour-coded as in Fig. 1d with the catalytic carboxylates shown. The AC74 duplex RNA is shown in brown with the primer terminal C74 and the incoming CTP in full atom colouring. Pol β DNA and the incoming dCTP are in light blue. The anomalous difference electron densities (3σ) arising from the Mn²⁺ ions

(cyan) are shown in green. **b**, Different orientations of the primer duplexes in class I (red) and class II (yellow) CCA-adding enzymes, and in Pol β (blue). The surface representation of the class II CCA-adding enzyme is shown with the modelled tRNA acceptor stem and TΨC stem loop¹⁰. The primer duplexes in the structures of AfCCA and Pol β ternary complexes are aligned onto the class II BstCCA enzyme by superimposing their homologous head domains (magenta).

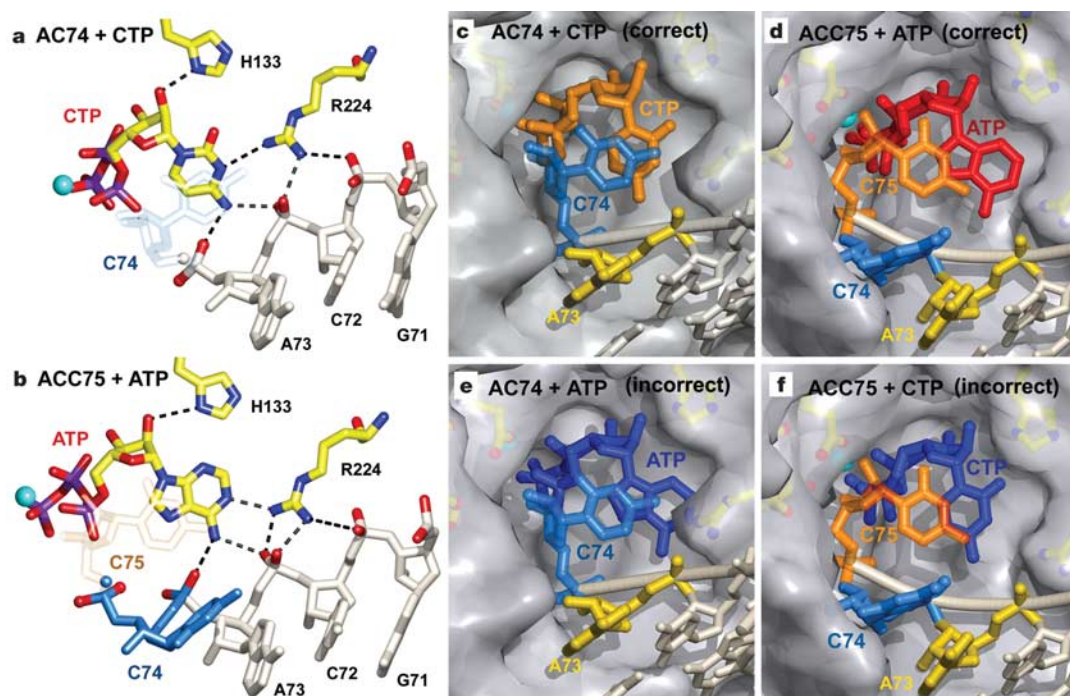


Figure 4 Nucleotide specificity. **a**, **b**, Hydrogen-bonding complementarity between the base of the incoming NTP and the nucleotide-binding site. The incoming NTPs and selected protein side chains are in full atom colouring with hydrogen-bonding interactions represented by dashed lines. The nucleotides C74 and C75 are shown in blue and orange, respectively, whereas other nucleotides are in grey. The cyan sphere represents metal ion B. **c–f**, Changes in the size of the nucleotide-binding pocket at the last two stages of nucleotide addition. The enzyme is shown in a semi-transparent surface representation

with selected side chains shown underneath. The colouring of the nucleotides is the same as in Fig. 2. **c**, The AC74 complex with the observed incoming CTP (orange). **d**, The ACC75 complex with observed incoming ATP (red). **e**, The AC74 complex with an incorrect incoming ATP (blue) modelled by superimposing its triphosphate on that of the observed CTP in **c**. **f**, The ACC75 complex with an incorrect incoming CTP (blue) modelled by superimposing its triphosphate on that of the observed ATP in **d**.

conserved between class I CCA-adding enzymes, adopts different rotamers depending on whether or not the tRNA is present. Thus, the tRNA and the protein collaborate to form a binding pocket for the incoming nucleotide, consistent in some ways with an earlier proposal based on biochemical studies¹⁸.

Specificity for nucleotide selection and termination

It is clear from the hydrogen-bonding patterns between the incoming bases and the enzyme in the ternary complexes that discrimination against UTP and GTP arises owing to their lack of hydrogen-bonding complementarity to the binding site for the base (Fig. 4a, b). Because the backbone phosphate oxygens of the tRNA that form part of the base-recognition site are obligate hydrogen bond acceptors, they can only interact with hydrogen bond donors such as the N4 of CTP or the N6 of ATP. By the same token, base recognition by Arg 224 requires an acceptor such as the N3 of CTP or the N1 of ATP. As with the class II CCA-adding enzyme of *B. stearothermophilus*⁸, the Watson–Crick face of U nor G is not complementary to the base-binding site because their patterns of hydrogen donors and acceptors are the reverse of that for C and A. As a consequence, UTP and GTP are excluded from the binding site. The difference between the two classes of enzyme, however, seems to be whether the discrimination requires tRNA. With class II CCA-adding enzymes, the incoming NTPs can be recognized by protein side chains⁸, albeit in the absence of tRNA.

In contrast to its discrimination against UTP and GTP, the CCA-adding enzyme uses the more subtle strategy of exploiting the size difference between pyrimidine and purine bases to distinguish between CTP and ATP. The size of the nucleotide-binding pocket changes to accommodate the appropriate incoming NTP at different steps of nucleotide incorporation. For the tRNA analogue terminated at nucleotide C74, the binding pocket provides a snug fit for CTP (Fig. 4c), but is too small to allow the correct positioning

of both the α -phosphate and base of an ATP. If an ATP molecule enters the catalytic site with its triphosphate bound to the location required for catalytic insertion, its Watson–Crick base edge clashes with both the enzyme and the tRNA (Fig. 4d). After the addition of C75, the binding pocket becomes enlarged sufficiently to accommodate an ATP (Fig. 4e). Because the binding pocket can still accommodate CTP, discrimination against CTP at this step may not be absolute. In this enlarged binding site the base of a bound CTP cannot reach and interact with Arg 224 and tRNA specificity determinants, and thus presumably cannot achieve an enzymatically productive binding (Fig. 4f). Indeed, biochemical studies *in vitro* show that in the absence of ATP both classes of CCA-adding enzyme can add a poly(C) tail to the 3' terminus of immature tRNA¹⁵, with a twofold increase in K_m and a 20-fold decrease in k_{cat} for CTP compared with ATP at the step of adding nucleotide 76.

The interactions between the growing 3' terminus of the tRNA and the enzyme allow the enzyme to detect the length of the 3' terminus and change the size of the binding pocket. When C74 is presented to the active site as the primer nucleotide, the β -turn residues Ala 95 and Glu 96, which are part of the binding pocket opposite to the base-determinant side, stack onto A73 so that the enzyme is in a slightly closed conformation. This conformation produces a binding pocket that fits CTP. After C75 is added, C74 moves out of the primer position and makes extensive hydrogen-bonding interactions with protein residues 95–99. These interactions cause unstacking between the enzyme and A73, and drive open the nucleotide-binding site into a conformation that accommodates ATP. Because these interactions with C74 make use of all three Watson–Crick hydrogen-bonding sites of the base C, they are highly specific and might potentially provide a proofreading mechanism that detects an incorrectly incorporated nucleotide at position 74. Only the 3' terminus with a correctly incorporated C74 would be able to alter the binding pocket and proceed to the addition of A76.

The termination of nucleotide addition seems to be associated with the formation of a stable, helically stacked tRNA CCA end, the 3' terminus of which is not positioned in the primer-binding site. The stacked 3'-terminal nucleotides of the tRNA product are packed against a β -sheet in the catalytic site of the AfCCA complex (Fig. 2c), which may further stabilize the 3'-terminal nucleotides in the stacked conformation. The stability of this stacked base conformation may prevent A76 from adopting a catalytically active geometry for the next round of nucleotide addition, which would require that A76 rotate about 90° away from its observed position. Furthermore, the limited binding pocket size seems to be insufficient to accommodate the two-nucleotide bulge required to position A76 for further incorporation.

The requirement for a movable head domain

The ability of the flexibly attached catalytic head domain to reposition in response to the length of the growing 3' end of the tRNA is responsible for the alterations of the binding pocket that render it specific for CTP, ATP or termination. The changes in the size and shape of the binding site are not generated by local structural rearrangements, but rather they are achieved through a progressive repositioning of the head domain (Fig. 5). The flexibility of the head domain seems to be an intrinsic property of the AfCCA enzyme, as it has a high thermal B -factor and its orientation relative to the rest of the enzyme differs slightly in all of the complex structures. In fact, the head domain was largely disordered in one of the *apo*-AfCCA crystal forms¹⁰. The growing 3' terminus of tRNA—presumably with its 3'-terminal nucleotide in an equilibrium between stacked and flipped-out conformations—along with the incoming NTP locks down a conformation of the flexible head domain that produces a pocket size appropriate for the correct incoming nucleotide at a given addition step, or for the final stacked structure required for the termination of the reaction. On the basis

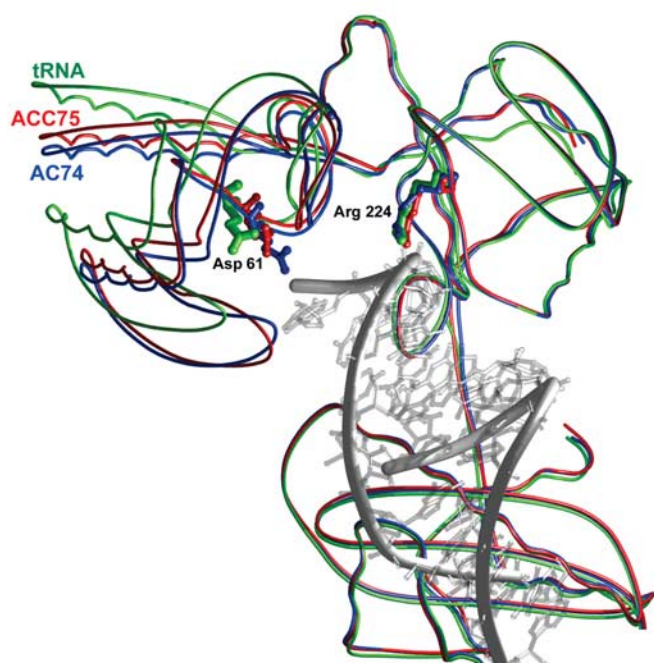


Figure 5 Movement of the head domain. AfCCA enzymes from the AC74 (blue), the ACC75 (red) and the tRNA (green) complexes are aligned through superimposing their tRNA-binding domains. The ACC75 RNA is shown in grey. Arg 224 and Asp 61, which are residues situated at opposite sides of the nucleotide-binding pocket, are shown.

of our modelling studies, it seems that the ability of the head domain to reposition may also be required for the addition of the first C in the CCA triplet.

Discussion

The mechanism by which the class I CCA-adding enzymes are able to specifically add CCA to the end of immature tRNA without a nucleic acid template contains aspects of previous proposals plus numerous unanticipated features. The growing primer terminus is scrunched in the active site as it elongates, and both the tRNA and protein cooperate to provide a template for the incoming base. The enzyme's specificity for adding a C or A, or terminating, however, arises from changes in the size of the nucleotide-binding site, which is altered by the elongating tRNA (Supplementary Movie 2).

Note added in proof: The crystal structure of a class II CCA-adding enzyme complex with tRNA presented in the accompanying paper²⁷ agrees with our modelling of the class II acceptor stem duplex and priming terminal nucleotide orientations shown in Fig. 3. □

Methods

Crystallization and data collection

The AfCCA molecule containing a his-tag at its C terminus was expressed and purified as described previously¹⁶. The tRNA^{phe} from baker's yeast was purchased from Sigma and the synthetic oligonucleotides AC74 and ACC75 were purchased from Dharmacon. AfCCA (5 mg ml⁻¹) and tRNA^{phe} were mixed at a molar ratio of 1:1.5 in a sample buffer containing 20 mM Tris (pH 7.5), 5 mM MgCl₂ and 150 mM NaCl. Crystals grew by the hanging-drop vapour diffusion method after six months at 4 °C from a 1:1 mixture of sample (2 µl) with a solution containing 1.5 M (NH₄)₂SO₄, 100 mM Tris (pH 8.5) and 12% glycerol. The crystals diffracted to 6.5 Å resolution at synchrotron beamlines.

The duplex RNA oligonucleotides were mixed with AfCCA (5 mg ml⁻¹) at a molar ratio of 3:1 in the same sample buffer used for the tRNA complex. Crystals of AC74 and ACC75 were obtained after two weeks at room temperature in 70 mM sodium tartrate and 7% PEG 3350. Crystals complexed with NTPs were obtained by soaking the binary crystals in mother liquor containing 4 mM ATP or CTP. To trap the complex in a catalytically active conformation but avoid nucleotide incorporation in the crystal, all soaking experiments were carried out at 4 °C for about 2 h. Furthermore, the reaction was inhibited by the tartrate buffer, which chelated the Mg²⁺ ions and prevented their binding to the metal position A, as determined by the crystal structures. To observe the binding of metal ion A, crystals of the AC74 ternary complex were soaked in a buffer containing chloride and Mn²⁺, instead of tartrate and Mg²⁺, for 30 min at 4 °C. Our attempt to repeat this procedure with the ACC75 complex was unsuccessful due to the limited number of crystals. The ACC75 ternary complex was also obtained by co-crystallization with a non-reactive αβ-methylene ATP. Data for the ternary complex crystals were collected at resolutions from 2.2 Å to 3.4 Å and were processed using the HKL package²⁰. The statistics are summarized in Table 1.

Structure determination and refinement

The structure of AfCCA complexed with tRNA was solved by molecular replacement with the program AMoRe²¹ using the apo-AfCCA dimer structure as a search model. The program RESOLVE²² was used for density modification and two-fold non-crystallographic symmetry (NCS) averaging. Rigid-body refinement of the model was carried out using the program Refmac²³. The ternary complexes were also solved by molecular replacement with AMoRe using the same search model. The crystallographic asymmetric unit contains two AfCCA dimers bound to three duplex RNAs. One of the duplex RNA molecules is shared by two AfCCA proteins, which presumably is a crystallization feature. The electron density maps calculated using model-derived phases were improved by cycles of multi-domain cross crystal averaging using the program Dmmulti²⁴ and real-space NCS averaging within the crystal using the program RAVE²⁵. The final model was rebuilt with the graphics program O²⁶ and refined with Refmac.

Received 4 April; accepted 3 June 2004; doi:10.1038/nature02711.

1. Deutscher, M. P. tRNA nucleotidyltransferase. *The Enzymes* **15**, 183–215 (1982).
2. Green, R. & Noller, H. F. Ribosomes and translation. *Annu. Rev. Biochem.* **66**, 679–716 (1997).

3. Nissen, P., Hansen, J., Ban, N., Moore, P. B. & Steitz, T. A. The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**, 920–930 (2000).
4. Aebi, M. *et al.* Isolation of a temperature-sensitive mutant with an altered tRNA nucleotidyltransferase and cloning of the gene encoding tRNA nucleotidyltransferase in the yeast *Saccharomyces cerevisiae*. *J. Biol. Chem.* **265**, 16216–16220 (1990).
5. Zhu, L. & Deutscher, M. P. tRNA nucleotidyltransferase is not essential for *Escherichia coli* viability. *EMBO J.* **6**, 2473–2477 (1987).
6. Holm, L. & Sander, C. DNA polymerase β belongs to an ancient nucleotidyltransferase superfamily. *Trends Biochem. Sci.* **20**, 345–347 (1995).
7. Yue, D., Maizels, N. & Weiner, A. M. CCA-adding enzymes and poly(A) polymerases are all members of the same nucleotidyltransferase superfamily: characterization of the CCA-adding enzyme from the archaeal hyperthermophile *Sulfolobus shibatae*. *RNA* **2**, 895–908 (1996).
8. Li, F. *et al.* Crystal structures of the *Bacillus stearothermophilus* CCA-adding enzyme and its complexes with ATP or CTP. *Cell* **111**, 815–824 (2002).
9. Augustin, M. A. *et al.* Crystal structure of the human CCA-adding enzyme: insights into template-independent polymerization. *J. Mol. Biol.* **328**, 985–994 (2003).
10. Xiong, Y., Li, F., Wang, J., Weiner, A. M. & Steitz, T. A. Crystal structures of an archaeal class I CCA-adding enzyme and its nucleotide complexes. *Mol. Cell* **12**, 1165–1172 (2003).
11. Okabe, M. *et al.* Divergent evolutions of trinucleotide polymerization revealed by an archaeal CCA-adding enzyme structure. *EMBO J.* **22**, 5918–5927 (2003).
12. Steitz, T. A. A mechanism for all polymerases. *Nature* **391**, 231–232 (1998).
13. Deutscher, M. P. Reactions at the 3' terminus of transfer ribonucleic acid. 3. Catalytic properties of two purified rabbit liver transfer ribonucleic acid nucleotidyl transferases. *J. Biol. Chem.* **247**, 459–468 (1972).
14. Tomari, Y., Suzuki, T., Watanabe, K. & Ueda, T. The role of tightly bound ATP in *Escherichia coli* tRNA nucleotidyltransferase. *Genes Cells* **5**, 689–698 (2000).
15. Seth, M., Thurlow, D. L. & Hou, Y. M. Poly(C) synthesis by class I and class II CCA-adding enzymes. *Biochemistry* **41**, 4521–4532 (2002).
16. Yue, D., Weiner, A. M. & Maizels, N. The CCA-adding enzyme has a single active site. *J. Biol. Chem.* **273**, 29693–29700 (1998).
17. Shi, P. Y., Maizels, N. & Weiner, A. M. CCA addition by tRNA nucleotidyltransferase: polymerization without translocation? *EMBO J.* **17**, 3197–3206 (1998).
18. Shi, P. Y., Weiner, A. M. & Maizels, N. A top-half tDNA minihelix is a good substrate for the eubacterial CCA-adding enzyme. *RNA* **4**, 276–284 (1998).
19. Pelletier, H., Sawaya, M. R., Kumar, A., Wilson, S. H. & Kraut, J. Structures of ternary complexes of rat DNA polymerase beta, a DNA template-primer, and ddCTP. *Science* **264**, 1891–1903 (1994).
20. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
21. Navaza, J. Implementation of molecular replacement in AMoRe. *Acta Crystallogr. D* **57**, 1367–1372 (2001).
22. Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55**, 849–861 (1999).
23. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
24. Cowtan, K. 'dm': An automated procedure for phase improvement by density modification. *Joint CCP4 ESF-EACBM Newsl. Prot. Crystallogr.* **31**, 34–38 (1994).
25. Kleywegt, G. J. & Jones, T. A. Software for handling macromolecular envelopes. *Acta Crystallogr. D* **55**, 941–944 (1999).
26. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
27. Tomita, K. *et al.* Structural basis for template-independent RNA polymerization. *Nature* doi:10.1038/nature02712 (this issue).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank the beamline staff at X25 and X12C of the National Synchrotron Light Source, 8BM at the Advanced Photon Source, A1 at the Cornell High Energy Synchrotron Source, and 8.2.1 and 8.2.2 at the Advanced Light Source for assistance in data collection. We thank A. M. Weiner for suggestions on the manuscript; J. Wang for discussions; members of the Steitz laboratory for assistance at various stages of the project; and R. Evans for work with the ACC75 complex. This work was supported by an NIH grant.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.A.S. (eatherton@csb.yale.edu). The atomic coordinates and diffraction amplitudes have been deposited in the Protein Data Bank under accession codes 1SZ1, 1TFY and 1TFW for the tRNA, AC74 and ACC75 complexes, respectively.

An apparently normal γ -ray burst with an unusually low luminosity

S. Yu. Sazonov^{1,2}, A. A. Lutovinov¹ & R. A. Sunyaev^{1,2}

¹Space Research Institute, Russian Academy of Sciences, Profsoyuznaya 84/32, 117997 Moscow, Russia

²Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, D-85740 Garching bei München, Germany

Much of the progress in understanding γ -ray bursts (GRBs) has come from studies of distant events (redshift $z \approx 1$). In the brightest GRBs, the γ -rays are so highly collimated that the events can be seen across the Universe. It has long been suspected that the nearest and most common events have been missed because they are not as collimated or they are under-energetic (or both)¹. Here we report soft γ -ray observations of GRB 031203, the nearest event to date ($z = 0.106$; ref. 2). It had a duration of 40 s and peak energy of >190 keV, and therefore appears to be a typical long-duration GRB. The isotropic γ -ray energy of $\leq 10^{50}$ erg, however, is about three orders of magnitude smaller than that of the cosmological population. This event—as well as the other nearby but somewhat controversial GRB 980425—is a clear outlier from the isotropic-energy/peak-energy relation^{3,4} and luminosity/spectral-lag relations^{5,6} that describe the majority of GRBs. Radio calorimetry shows that both of these events are under-energetic explosions⁷. We conclude that there does indeed exist a large population of under-energetic events.

On 2003 December 3 at 22:01:28 UTC, IBIS, a hard X-ray coded aperture mask imager on the INTEGRAL satellite, detected⁸ a pulse of 40 s duration. This event, GRB 031203, was localized by on-board software and the 2.5-arcmin position rapidly disseminated⁹. The event, with a simple profile (Fig. 1), appears to be a typical long-duration GRB. Likewise the spectrum is also typical (Fig. 2). A single power law model with photon index $\alpha = -1.63 \pm 0.06$ provides an adequate fit. We place a lower limit on the spectral peak energy, $E_{\text{peak}} > 190$ keV (90% confidence; see Fig. 2).

We found no evidence for significant spectral evolution on short (seconds) timescales. Next, we cross-correlated the light curves in two energy ranges (soft, 20–50 keV, and hard, 100–200 keV) and detected a marginal lag of 0.24 ± 0.12 in the usual sense (harder emission preceding the softer emission).

Watson *et al.*¹⁰ have suggested that GRB 031203 is an X-ray flash (XRF). XRFs are defined^{11,12} either by $E_{\text{peak}} < 50$ keV or by a larger 2–30 keV fluence (S_X) compared to that in the traditional GRB band, 30–400 keV (S_γ). The high- E_{peak} soft γ -ray spectrum measured by INTEGRAL provides direct evidence that GRB 031203 is a GRB rather than an XRF. The case for an XRF was made by a very high value of flux at 1 keV, $F_X = (2.6 \pm 1.3) \times 10^{-6} \text{ erg cm}^{-2} \text{ keV}^{-1}$, inferred from modelling of a dust-scattered echo¹³. As can be seen from Fig. 2 the soft X-ray emission is well above an extrapolation of the INTEGRAL spectrum (which predicts $0.36 < S_X/S_\gamma < 0.53$, depending on the precise value of E_{peak}) and thus it must have an origin different from the process producing the soft γ -ray pulse. It is possible that the echo was caused by the early ($\leq 1,000$ s) afterglow. We note however that the inferred soft X-ray fluence is inversely proportional to the assumed dust column and extrapolation of our spectrum to keV energies is consistent with the lower soft X-ray fluence (due to a higher dust column) advocated by Prochaska *et al.*².

The burst fluence in the 20–200 keV band is $(2.0 \pm 0.4) \times 10^{-6} \text{ erg cm}^{-2}$. Adopting the redshift of 0.1 (ref. 2) and the currently popular cosmological parameters $(H_0, \Omega_m, \Omega_\Lambda) = (75 \text{ km s}^{-1} \text{ Mpc}^{-1}, 0.3, 0.7)$, we find that the isotropic energy equivalent is $(4 \pm 1) \times 10^{49} \text{ erg}$. Defining $\epsilon_{\gamma, \text{iso}}$ to be the isotropic

energy equivalent over the 20–2000 keV band, we find $6 \times 10^{49} \text{ erg} < \epsilon_{\gamma, \text{iso}} < 1.4 \times 10^{50} \text{ erg}$; where the range reflects the observational uncertainty in the spectrum above 200 keV (see Fig. 2).

GRB 031203, an event with spectrum similar to cosmological GRBs, is the least energetic (in terms of $\epsilon_{\gamma, \text{iso}}$) long-duration GRB. Clearly, γ -ray luminosities and energy releases vary widely, spanning at least four orders of magnitude. Furthermore it is of considerable interest to note that GRB 031203 violates two much discussed relations in GRB astrophysics: (1) the $\epsilon_{\gamma, \text{iso}}-E_{\text{peak}}$ relation and (2) the luminosity–spectral lag relation. For GRB 031203, the first relation predicts⁴ $E_{\text{peak}} \approx 10$ keV, in gross disagreement with the analysis presented here. Long spectral lags are expected from the second relation^{5,6} when in fact we see virtually no lag (Fig. 3).

GRB 031203, however, shares some properties with GRB 980425, which is associated with a nearby ($z = 0.0085$) supernova, SN 1998bw (refs 14, 15). This event was also severely underenergetic, $\epsilon_{\gamma, \text{iso}} \approx 10^{48} \text{ erg}$, and violated the $\epsilon_{\gamma, \text{iso}}-E_{\text{peak}}$ relation. Curiously, GRB 98425 was also a single pulse¹⁶ but without a cusp.

To summarize, the two nearest long-duration events, GRB 031203 and GRB 980425, are clearly sub-energetic in the γ -ray band. Their proximity (and hence implied abundance) makes it of great interest to understand their origin and relation to the more distant cosmological events. Are these events genuinely low-energy explosions¹⁵ ('sub-energetic' model) or a typical GRB viewed away from its axis ('off-axis' model)?

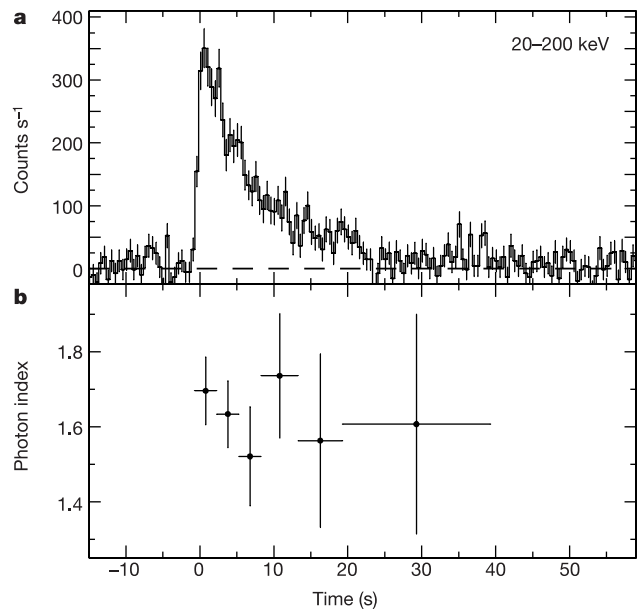


Figure 1 The temporal profile of GRB 031203 and its evolution. **a**, The profile in the 20–200 keV energy range obtained with the IBIS/ISGRI detector on board INTEGRAL. The binning is 0.5 s. Time is measured relative to burst trigger. A background level (136 counts s^{-1}) was estimated from a 200-s interval preceding the trigger and then subtracted from the profile. Vertical error bars indicate poissonian noise. The profile can be classified as a FRED ('fast rise exponential decay') with a rise time of about 1 s and an e -folding time of 8 ± 0.5 s. The peak flux is $2.6 \text{ photons cm}^{-2} \text{ s}^{-1}$ corresponding to $2.4 \times 10^{-7} \text{ erg cm}^{-2} \text{ s}^{-1}$ in the 20–200 keV band. Two X-ray sources present in the field of view (Vela X-1 and 4U 0836–429) contribute only $\sim 15 \text{ counts s}^{-1}$ to the total count rate (before background subtraction). Imaging analysis of IBIS data revealed no source at the position of GRB 031203 during half an hour before, or one day after, the burst. The corresponding 3σ upper limits (in the 20–200 keV band) are $\sim 10^{-9} \text{ erg cm}^{-2} \text{ s}^{-1}$ and $\sim 10^{-10} \text{ erg cm}^{-2} \text{ s}^{-1}$, respectively. **b**, The evolution of the photon index during the duration of the burst. The spectrum over 20–200 keV is fitted to a single power law with index α ; the bin width varies from 2.5 s near the burst peak to 20 s during the decay phase. Vertical bars indicate 1σ uncertainties.

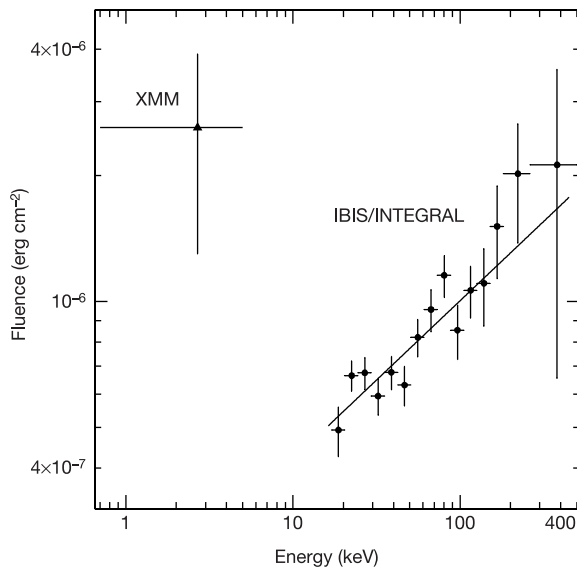


Figure 2 Spectral energy distribution of GRB 031203 shown in νF_ν units. The data points in the 17–500 keV range were obtained from the data of the IBIS/ISGRI detector for the first 20 s of the burst, when 80% of its total energy was emitted. Scattering and absorption in the interstellar medium of our Galaxy and the host galaxy of GRB 031203 has negligible effect ($<1\%$) on observed flux at photon energies above 20 keV. The energy bin width varies from 3 keV at 17 keV to 250 keV at 500 keV. Vertical bars indicate 1σ statistical uncertainties. We considered a single power law model (photon index, α). Our method^{23,24} consisted of constructing images in predefined energy intervals followed by normalizing the resulting source fluxes to the corresponding fluxes of the Crab nebula for a similar position in the field of view. Analysis of an extensive set of Crab calibration observations has shown that the source absolute flux can be recovered with an accuracy of 10% and the systematic uncertainty of relative flux measurement in different energy channels is less than 5%. The latter uncertainty was included in the modelling of the spectrum. The best-fit power law model with $\alpha = -1.63 \pm 0.06$ (1σ uncertainty, $\chi^2/\text{d.o.f.} = 14.8/15$) is shown by the line. We also considered a double power law model (the so-called ‘Band’ model^{25,26}). Setting the high-energy power law index to -2 we are able to place a lower limit to the peak energy, $E_{\text{peak}} > 190$ keV (90% confidence level). The data point towards the top left corner of the figure is the soft X-ray (0.7–5 keV) fluence, F_x , inferred^{10,13} from the dust scattered haloes discovered in XMM-Newton observations. The vertical bars indicate 1σ uncertainties.

In the off-axis model¹⁷, $\epsilon_{\gamma, \text{iso}} \propto \delta^n$ with $n \approx 2-3$, where $\delta = \gamma^{-1}[1 - \beta \cos(\theta - \theta_j)]^{-1}$ is the so-called Doppler factor; here, $v = c\beta$ is the velocity of the shocked ejecta, c is the velocity of light, $\gamma = (1 - \beta^2)^{-1/2}$, θ is the angle between the observer and the principal axis of the explosion and θ_j is the opening angle of the explosion (‘jet’). If we wish to make GRB 031203 to have isotropic energies similar to cosmological GRBs then δ should be $\sim 10-30$ times smaller than the on-axis value δ_0 . The true peak energy is then $(\delta_0/\delta) \times (1+z) \times E_{\text{peak}} > 2$ MeV — making GRB 031203 one of the hardest bursts. A second consequence is that the afterglow should brighten as the ejecta slows down. Soderberg *et al.*⁷ do not see any rebrightening, and furthermore the afterglow is faint, indicating that the explosion was under-energetic, $\lesssim 10^{50}$ erg. Likewise, radio calorimetry of SN 1998bw at early¹⁵ or late¹⁸ times finds $\lesssim 10^{50}$ erg. Thus we conclude that GRB 031203 and GRB 980425 are intrinsically sub-energetic events.

With a peak count rate of 1.3 photons $\text{cm}^{-2} \text{s}^{-1}$ (50–300 keV), GRB 031203 could have been detected by BATSE out to $z \approx 0.25$. So as not to significantly distort the observed (nearly flat) burst intensity distribution at low fluxes, less than ~ 300 underluminous bursts like GRB 031203 can be present in the BATSE catalogue¹⁹, including up to ~ 20 located at $z < 0.1$. On the other hand, a ‘typical’ GRB with $\epsilon_{\gamma, \text{iso}} \approx 10^{53}$ erg would have a fluence of 10^{-3} erg cm^{-2} if it occurred as close as GRB 031203. Only a few

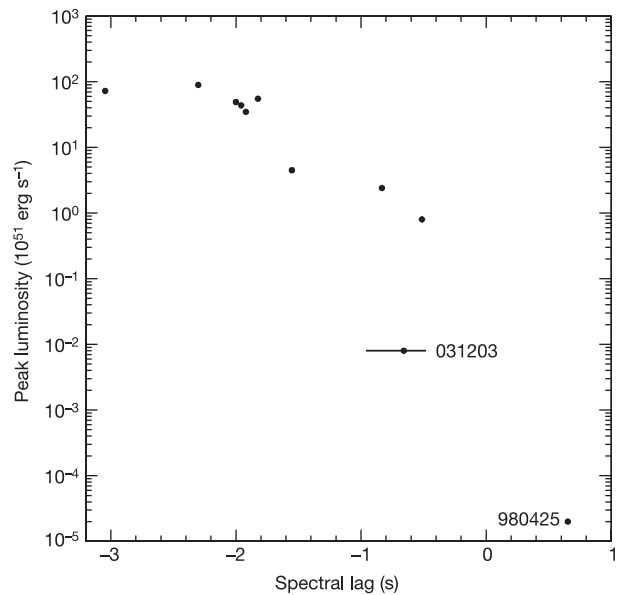


Figure 3 Spectral lag versus luminosity for cosmological and low-redshift GRBs. The data for cosmological bursts (unlabelled points), with measured redshifts ranging between $z = 0.84$ and 4.5 , are adopted from ref. 27. The lag is defined between the burst profiles at 25–50 keV and 100–300 keV, and the luminosity is calculated over the 50–300 keV band at the burst peak, assuming isotropic emission. Also plotted are the data for GRB 980425 ($z = 0.0085$)⁶, and GRB 031203 ($z = 0.1055$, this work). In the latter case, the peak luminosity measured with INTEGRAL was converted to the 50–300 keV range, and the lag (shown with 1σ error bars) was determined between the 20–50 keV and 100–200 keV bands. All the lags have been corrected for cosmological time dilation. The luminosities are given for a cosmology with $(H_0, \Omega_m, \Omega_\Lambda) = (65 \text{ km s}^{-1} \text{ Mpc}^{-1}, 0.3, 0.7)$.

such bright bursts have been observed in the ~ 30 years of GRB observations^{20–22}, suggesting a large population of events like GRB 031203 (see also ref. 2). We eagerly await the launch of the Swift mission, which with its increased sensitivity should detect and localize many ($\times 10$ the current rate) under-energetic events like GRB 031203. $\square$

Received 25 March; accepted 15 June 2004; doi:10.1038/nature02748.

- MacFadyen, A. I., Woosley, S. E. & Heger, A. Supernovae, jets and collapsars. *Astrophys. J.* **550**, 410–425 (2001).
- Prochaska, J. X. *et al.* The host galaxy of GRB 031203: implications of its low metallicity, low redshift, and starburst nature. *Astrophys. J.* (in the press).
- Amati, L. *et al.* Intrinsic spectra and energetics of BeppoSAX gamma-ray bursts with known redshifts. *Astron. Astrophys.* **390**, 81–89 (2002).
- Lamb, D. Q., Donaghy, T. Q. & Graziani, C. A unified jet model of X-ray flashes and γ -ray bursts. *New Astron. Rev.* **48**, 459–464 (2004).
- Norris, J. P., Marani, G. F. & Bonnell, J. T. Connection between energy-dependent lags and peak luminosity in gamma-ray bursts. *Astrophys. J.* **534**, 248–257 (2000).
- Schaefer, B. E., Deng, M. & Band, D. L. Redshifts and luminosities for 112 gamma-ray bursts. *Astrophys. J.* **563**, L123–L127 (2001).
- Soderberg, A. M. *et al.* The sub-energetic γ -ray burst GRB 031203 as a cosmic analogue to the nearby GRB 980425. *Nature* doi:10.1038/nature02757 (this issue).
- Mereghetti, S. & Götz, D. GRB 031203: further analysis of INTEGRAL data. *GCN Circ.*, 2460 (2003).
- Götz, D., Mereghetti, S., Beck, M., Borkowski, J. & Mowlavi, N. GRB 031203: a long GRB detected with INTEGRAL. *GCN Circ.* 2459 (2003).
- Watson, D. *et al.* A very low luminosity X-ray flash: XMM-Newton observations of GRB 031203. *Astrophys. J.* **605**, L101–L104 (2004).
- Heise, J., in’t Zand, J., Kippen, R. M. & Woods, P. M. in *Gamma-ray Bursts in the Afterglow Era* (eds Costa, E., Frontera, F. & Hjorth, J.) 16–21 (Springer, Berlin/Heidelberg, 2001).
- Barraud, C. *et al.* Spectral analysis of 35 GRBs/XRFs observed with HETE-2/FREGATE. *Astron. Astrophys.* **400**, 1021–1030 (2003).
- Vaughan, S. *et al.* The discovery of an evolving dust-scattered X-ray halo around GRB 031203. *Astrophys. J.* **603**, L5–L8 (2004).
- Galama, T. J. *et al.* An unusual supernova in the error box of the gamma-ray burst of 25 April 1998. *Nature* **395**, 670–672 (1998).
- Kulkarni, S. R. *et al.* Radio emission from the unusual supernova 1998bw and its association with the gamma-ray burst of 25 April 1998. *Nature* **395**, 663–669 (1998).
- Bloom, J. S. *et al.* Expected characteristics of the subclass of supernova gamma-ray bursts. *Astrophys. J.* **506**, L105–L108 (1998).

17. Yamazaki, R., Yonetoku, D. & Nakamura, T. An off-axis jet model for GRB 980425 and low-energy gamma-ray bursts. *Astrophys. J.* **594**, L79–L82 (2003).
18. Soderberg, A. M., Frail, D. A. & Wieringa, M. H. Constraints on off-axis gamma-ray burst jets in type Ibc supernovae from late-time radio observations. *Astrophys. J.* **607**, L13–L16 (2004).
19. Paciesas, W. S. *et al.* The fourth BATSE gamma-ray burst catalog (revised). *Astrophys. J. Suppl. Ser.* **122**, 465–495 (1999).
20. Mazets, E. P. & Golenetskii, S. V. Recent results from the gamma-ray burst studies in the KONUS experiment. *Astrophys. Space Sci.* **75**, 47–81 (1981).
21. Kuznetsov, A. V. *et al.* SIGNE:2 MP9 data for the powerful gamma-ray burst of 1983 Aug 1. *Sov. Astron. Lett.* **12**, 315–318 (1986).
22. Fenimore, E. E. *et al.* The intrinsic luminosity of gamma-ray bursts and their host galaxies. *Nature* **366**, 40–42 (1993).
23. Lutovinov, A. A., Molkov, S. V. & Revnivtsev, M. G. The first results of observations of the transient pulsar SAX J2103.5+4545 by the INTEGRAL observatory. *Astron. Lett.* **29**, 713–718 (2003).
24. Revnivtsev, M. G. *et al.* A hard X-ray survey of the Galactic-Center region with the IBIS telescope of the INTEGRAL observatory: a catalog of sources. *Astron. Lett.* **30**, 430–435 (2004).
25. Mazets, E. P. *et al.* Cosmic gamma-ray burst spectroscopy. *Astrophys. Space Sci.* **82**, 261–282 (1982).
26. Band, D. *et al.* BATSE observations of gamma-ray burst spectra. I - Spectral diversity. *Astrophys. J.* **413**, 281–292 (1993).
27. Schaefer, B. E. Gamma-ray burst Hubble diagram to $z = 4.5$. *Astrophys. J.* **583**, L67–L70 (2003).

Acknowledgements This work is based on a Core Programme pointed observation (PI: S.Yu.S.) with INTEGRAL, an ESA project with instruments and science data centre funded by ESA member states (especially the PI countries: Denmark, France, Germany, Italy, Switzerland, Spain), Czech Republic and Poland, and with the participation of Russia and the USA. We thank M. Revnivtsev and E. Churazov for help in the data analysis, and S. Kulkarni for suggestions.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.Yu.S. (sazonov@mpa-garching.mpg.de).

The sub-energetic γ -ray burst GRB 031203 as a cosmic analogue to the nearby GRB 980425

A. M. Soderberg¹, S. R. Kulkarni¹, E. Berger¹, D. W. Fox¹, M. Sako², D. A. Frail³, A. Gal-Yam¹, D. S. Moon⁴, S. B. Cenko⁴, S. A. Yost⁴, M. M. Phillips⁵, S. E. Persson⁵, W. L. Freedman⁵, P. Wyatt⁵, R. Jayawardhana⁶ & D. Paulson⁶

¹Caltech Optical Observatories 105-24, California Institute of Technology, Pasadena, California 91125, USA

²Stanford Linear Accelerator Center, 2575 Sand Hill Road M/S 29, Menlo Park, California 94025, USA

³National Radio Astronomy Observatory, PO Box 0, Socorro, New Mexico 87801, USA

⁴Space Radiation Laboratory 220-47, California Institute of Technology, Pasadena, California 91125, USA

⁵Carnegie Observatories, 813 Santa Barbara Street, Pasadena, California 91101, USA

⁶Department of Astronomy, University of Michigan, 830 Dennison Bldg, Ann Arbor, Michigan 48109, USA

Over the six years since the discovery¹ of the γ -ray burst GRB 980425, which was associated² with the nearby (distance ~ 40 Mpc) supernova 1998bw, astronomers have debated fiercely the nature of this event. Relative to bursts located at cosmological distance (redshift $z \approx 1$), GRB 980425 was under-luminous in γ -rays by three orders of magnitude. Radio calorimetry^{3,4} showed that the explosion was sub-energetic by a factor of 10. Here we report observations of the radio and X-ray afterglow of the recent GRB 031203 (refs 5–7), which has a redshift of $z = 0.105$. We demonstrate that it too is sub-energetic which, when taken together with the low γ -ray luminosity⁷, suggests that GRB 031203 is the first cosmic analogue to GRB 980425. We find no evidence that this event was a highly collimated explosion

viewed off-axis. Like GRB 980425, GRB 031203 appears to be an intrinsically sub-energetic γ -ray burst. Such sub-energetic events have faint afterglows. We expect intensive follow-up of faint bursts with smooth γ -ray light curves^{8,9} (common to both GRB 031203 and 980425) to reveal a large population of such events.

On 3 December 2003 at 22:01:28 UT, the INTEGRAL satellite detected^{5,7} a seemingly typical long-duration ($\Delta t \approx 20$ s) γ -ray burst. Within 6 h, the Newton X-ray Multiple Mirror (XMM) observatory detected^{10,11} an X-ray source with flux (2–10 keV band) $F_X = (3.95 \pm 0.09) \times 10^{-13}$ erg cm⁻² s⁻¹, fading gradually $\propto t^\alpha$ with $\alpha = -0.4$. Using the Very Large Array (VLA), we discovered a radio source at right ascension $\alpha(\text{J2000}) = 08 \text{ h } 02 \text{ min } 30.18 \text{ s}$ and declination $\delta(\text{J2000}) = -39^\circ 51' 03.51''$ (± 0.1 arcsec in each axis), well within the 6-arcsec radius error circle of the XMM source. A subsequent XMM observation¹² confirmed the gradual decay of the X-ray source. From our analysis of the XMM data, we find the flux $\propto t^{-0.4}$ between the two epochs and the spectral flux density, $F_{\nu X} \propto \nu^\beta$, is fitted by $\beta = -0.81 \pm 0.05$ with an absorbing column density, $N_H = 6.2 \times 10^{21}$ cm⁻². Taken

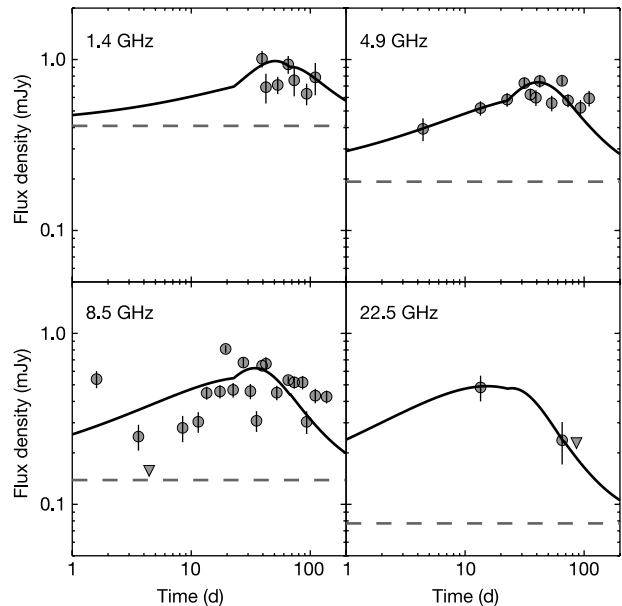


Figure 1 Radio light curves of the afterglow of GRB 031203. All measurements (circles) are summarized in Table 1 and include 1 σ error bars. Triangles represent 2 σ upper limits. The solid lines are models of synchrotron (afterglow) emission from spherical ejecta expanding into a uniform circumburst medium¹⁹. The models include a contribution from the host galaxy, which is well fitted by $F_{\text{host}} \approx 0.4(\nu/1.4 \text{ GHz})^{-0.6}$ mJy (dashed lines) and is consistent with the star-formation rate inferred⁶ from optical spectroscopy of the host. In applying the models, the X-ray observations are considered upper limits because they are probably dominated by (non-synchrotron) emission arising from the associated supernova SN 2003lw, as evidenced by the unusually slow flux decay at early time and the flat spectral index ($F_{\nu X} \propto t^{-0.4} \nu^{-0.8}$ as opposed to $\propto t^{-1} \nu^{-1.3}$ for GRBs). This was also the case for the X-ray emission¹ of GRB 980425/SN 1998bw ($F_{\nu X} \propto t^{-0.2} \nu^{-1}$). For our best-fit model, we find $\chi^2_r = 8.9$ (38 degrees of freedom), dominated by interstellar scintillation. The blastwave transitions to the non-relativistic regime at $t_{\text{NR}} \approx 23$ d. From the derived synchrotron parameters (at $t = 1$ d): $\nu_a \approx 3.2 \times 10^8$ Hz, $\nu_m \approx 3.6 \times 10^{12}$ Hz and $F_{\nu a} \approx 0.04$ mJy we find an isotropic afterglow energy, $E_{\text{AG,iso}} \approx 1.7 \times 10^{49} \nu_{c,15.5}^{1/4}$ erg, a circumburst density $n \approx 0.6 \nu_{c,15.5}^{3/4}$ cm⁻³ and the fractions of energy in the relativistic electrons (energy distribution $N(\gamma) \propto \gamma^{-p}$ with $p \approx 2.6$) and magnetic field of $\epsilon_e \approx 0.4 \nu_{c,15.5}^{1/4}$ and $\epsilon_B \approx 0.2 \nu_{c,15.5}^{-5/4}$, respectively. Here, $\nu_c = 3 \times 10^{15} \nu_{c,15.5}$ is the synchrotron cooling frequency, which is roughly constrained by the (non-synchrotron) SN 2003lw X-ray emission. Extrapolation of the synchrotron model beyond ν_c underestimates the observed X-ray flux by a factor of $\lesssim 10$, which is comparable to the discrepancy for SN 1998bw (found by extrapolating the radio model by Li and Chevalier⁴ ($p = 2.5$, $\epsilon_B = 10^{-3}$) beyond ν_c and comparing with the X-ray data¹ at $t \sim 12$ d).

Table 1 Radio observations made with the Very Large Array

Epoch (UT)	Δt (d)	$F_{1.43}$ (mJy)	$F_{4.86}$ (mJy)	$F_{8.46}$ (mJy)	$F_{22.5}$ (mJy)
2003 Dec. 5.52	1.60	—	—	0.540 ± 0.062	—
2003 Dec. 7.52	3.60	—	—	0.249 ± 0.043	—
2003 Dec. 8.35	4.43	—	0.393 ± 0.060	0.053 ± 0.052	—
2003 Dec. 12.38	8.46	—	—	0.280 ± 0.049	—
2003 Dec. 15.37	11.45	—	—	0.304 ± 0.042	—
2003 Dec. 17.38	13.46	—	0.520 ± 0.050	0.448 ± 0.039	0.483 ± 0.083
2003 Dec. 21.35	17.43	—	—	0.457 ± 0.041	—
2003 Dec. 23.37	19.45	—	—	0.811 ± 0.040	—
2003 Dec. 26.40	22.48	—	0.583 ± 0.054	0.467 ± 0.046	—
2003 Dec. 31.33	27.41	—	—	0.675 ± 0.045	—
2004 Jan. 4.33	31.41	—	0.728 ± 0.055	0.459 ± 0.047	—
2004 Jan. 8.26	35.34	—	0.624 ± 0.050	0.308 ± 0.043	—
2004 Jan. 12.29	39.37	1.011 ± 0.113	0.598 ± 0.063	0.647 ± 0.045	—
2004 Jan. 15.35	42.43	0.689 ± 0.136	0.749 ± 0.063	0.664 ± 0.061	—
2004 Jan. 25.24	52.32	0.710 ± 0.082	—	0.450 ± 0.044	—
2004 Jan. 26.34	53.42	—	0.556 ± 0.058	—	—
2004 Feb. 7.24	65.32	0.937 ± 0.112	0.751 ± 0.045	0.533 ± 0.028	0.273 ± 0.066
2004 Feb. 15.22	73.30	0.756 ± 0.147	0.576 ± 0.050	0.517 ± 0.042	—
2004 Feb. 28.13	86.21	—	—	0.517 ± 0.047	0 ± 0.114
2004 Mar. 6.17	93.25	0.631 ± 0.091	0.522 ± 0.058	0.304 ± 0.046	—
2004 Mar. 23.13	110.21	0.787 ± 0.169	0.593 ± 0.062	0.432 ± 0.042	—
2004 Apr. 19.07	137.15	—	—	0.426 ± 0.037	—

Observations commenced on 5 December 2003 UT. For all observations, we used the standard continuum mode with 2×50 MHz bands. At 22.5 GHz we used referenced pointing scans to correct for the systematic 10–20 arcsec pointing errors of the Very Large Array (VLA) antennas. We used the extra-galactic sources 3C 147 (J0542+498) and 3C 286 (J1331+305) for flux calibration, while the phase was monitored using J0828–375. The data were reduced and analysed using the Astronomical Image Processing System. The flux density and 1σ uncertainty were measured from the resulting maps by fitting a gaussian model to the afterglow emission.

together, the transient X-ray and radio emission are suggestive of afterglow emission.

In addition to monitoring the afterglow in various radio bands (Table 1 and discussion below) we obtained an observation of the source with the Advanced CCD Imaging Spectrometer (ACIS) instrument on board the Chandra X-ray observatory. The Chandra observations began on 22 January 2004 at 21:35 UT and lasted about 22 ks. We detected a faint source, count rate in the 2–10 keV band of $5.6 \times 10^{-4} \text{ s}^{-1}$, at $\alpha(\text{J2000}) = 08 \text{ h } 02 \text{ min } 30.159 \text{ s}$ and $\delta(\text{J2000}) = -39^\circ 51' 03.51''$ (± 0.18 arcsec in each axis), precisely coincident with the VLA source. Using the XMM model parameters stated above, we obtain $F_X = 6.4 \times 10^{-15} \text{ erg cm}^{-2} \text{ s}^{-1}$, implying a faster decline ($\alpha = -1 \pm 0.1$) between the second XMM and Chandra observations.

The primary interest in this burst is that the radio and X-ray afterglows coincide at the sub-arcsecond level¹³ with a nearby ($z = 0.1055$) galaxy⁶, making it the nearest GRB with the exception of the peculiar GRB 980425². At this redshift, the isotropic γ -ray energy release is 10^2 times smaller⁷ than that of the nearest classical event GRB 030329 ($z = 0.169$)¹⁴ and yet a factor of 10^2 larger^{1,2} than that of GRB 980425.

The afterglow properties of GRB 031203 also seem to be intermediate between classical cosmological GRBs and GRB 980425: the isotropic X-ray luminosity of GRB 031203 at $t \approx 10 \text{ h}$ is $L_X = 9 \times 10^{42} \text{ erg cm}^{-2} \text{ s}^{-1}$, nearly 10^3 times fainter than that observed¹⁵ for classical GRBs but a factor of 10^2 brighter¹ than that of GRB 980425. In the centimetre band, the peak luminosity is $L_{\nu, 8.5 \text{ GHz}} \approx 10^{29} \text{ erg s}^{-1} \text{ Hz}^{-1}$, fainter¹⁶ by a factor of 10^2 than that of most radio afterglows but comparable³ to that of GRB 980425. As L_X and the peak radio luminosity of an afterglow can be used^{17,15} as rough proxies for the afterglow energy, the data suggest that GRBs 031203 and 980425 are sub-energetic in comparison with classical GRBs.

As a next step, we applied the simplest afterglow model^{18,19} (a spherical relativistic blastwave shocking a constant density circum-burst medium and accelerating relativistic electrons: the afterglow emission arises from synchrotron emission of shocked electrons) to the afterglow data and obtain a satisfactory fit (Fig. 1). On the timescales best probed by the radio data—days to months—we see no evidence for a collimated (jet) geometry commonly seen²⁰ in the afterglows of cosmological GRBs.

From our modelling we confirm that the blast wave is sub-

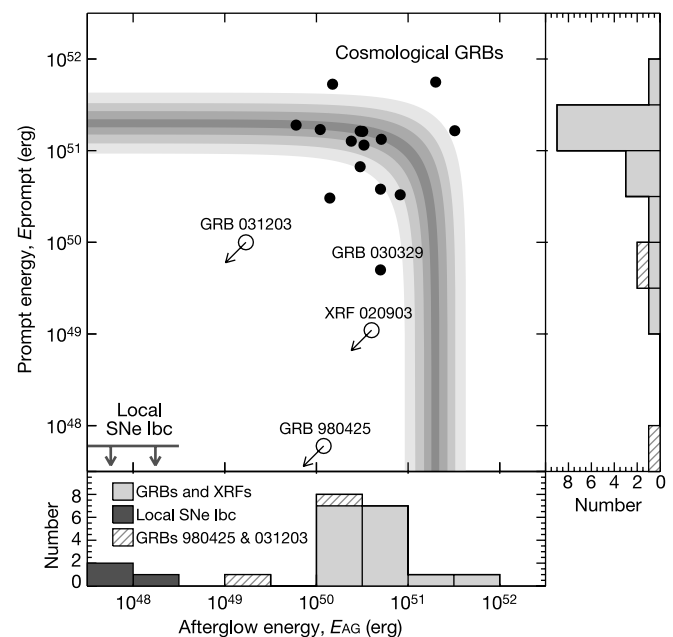


Figure 2 Two-dimensional energy plot for cosmic explosions. The energy in the prompt emission, E_{prompt} , and in the afterglow, E_{AG} , have been corrected^{20,26,27} for beaming based on the jet-break time observed for each burst, except in the cases of GRB 980425^{4,3}, X-ray flash (XRF) 020903²⁸ and GRB 031203 for which there is no evidence for a collimated outflow. For these three cases, we plot the isotropic values of E_{prompt} and E_{AG} and use an arrow to indicate they represent upper limits on both axes. The arcs mark lines of constant $E_{\text{prompt}} + E_{\text{AG}}$ as a guide to the reader. Most cosmological GRBs tend to cluster²⁷ around $E_{\text{prompt}} + E_{\text{AG}} \approx 2 \times 10^{51} \text{ erg}$, while GRBs 031203 and 980425, the nearest two bursts in the sample, are clearly sub-energetic. With the exception of SN 1998bw, associated with GRB 980425, there are no local Ibc supernovae with detected γ -ray emission, but the kinetic energy in the ejecta (excluding the photospheric energy yield) is generically found²⁹ to be $E_{\text{AG}} \lesssim 3 \times 10^{48} \text{ erg}$ (bottom left corner). Histograms of E_{AG} and E_{prompt} are shown in the bottom and side panels, respectively, for cosmological GRBs and local Ibc supernovae (SNe). The striped energy bins show the locations of GRBs 980425 and 031203.

energetic, finding an inferred afterglow energy of $E_{\text{AG}} \approx 1.7 \times 10^{49}$ erg. The circumburst particle density, $n \approx 0.6 \text{ cm}^{-3}$, is not atypical of that inferred¹⁷ for other GRBs. The blastwave is expected to become²¹ non-relativistic on a timescale $t_{\text{NR}} \approx 34(E_{\text{AG},50}/n_0)^{1/3} \text{ d}$, where we adopt the notation $q \equiv 10^x q_x$. The observational signatures²² of this transition, a steeper decay of the spectral peak frequency ($\nu_m \propto t^{-1.5} \rightarrow t^{-3}$) and an increase in the spectral peak flux ($F_{\nu_m} \propto t^0 \rightarrow t^{3/5}$) are consistent with the data (Fig. 1).

Here we use E_{AG} to denote the kinetic energy remaining in the blast wave after the prompt γ -ray energy release. In turn, the γ -ray emission arises from ultra-relativistic (bulk Lorentz factor $\Gamma \gtrsim 100$) ejecta within the blastwave. Thus, a more complete picture of the explosion energy is visualized through a two-dimensional plot of E_{prompt} , the beaming-corrected prompt energy release, versus E_{AG} (Fig. 2).

The two nearest events, GRBs 031203 and 980425, are clearly sub-energetic outliers in Fig. 2. Furthermore, we note several additional similarities: GRBs 031203 and 980425 (1) show no evidence for jets³; (2) possess simple γ -ray light curves^{1,7}, (3) violate⁷ the $E_{\text{prompt}} - E_{\text{peak}}$ relation²³ with respect to cosmological ('classical') bursts; and (4) are outliers in the luminosity-spectral lag relation⁹. This discussion motivates the question: How are these two events related to cosmological GRBs?

It has been suggested (see, for example, ref. 24) that all GRB explosions have the same energetics and explosion geometry. In this framework, sub-energetic bursts are simply events viewed away from the jet axis. Such bursts should have a soft E_{peak} and also exhibit a rise in the inferred E_{AG} as shocked ejecta eventually come into our line of sight. For GRB 031203, $E_{\text{peak}} > 190 \text{ keV}$ (ref. 7), comparable to cosmological GRBs for which we have observational evidence favouring an on-axis viewing angle. Moreover, we see no evidence for an increase in E_{AG} during the timescale of the radio observation ($\sim 150 \text{ d}$). Similarly, there is no evidence that E_{AG} is increasing for GRB 980425, despite dedicated radio monitoring²⁵ of the source since 1998. With no indication of their being off-axis explosions, we at present conclude that GRBs 031203 and 980425 are intrinsically sub-energetic events.

Astronomers have had to wait six years to discover a sub-energetic event similar to GRB 980425, despite a large population of such events (as implied by their proximity). The bulk of the population has escaped our attention owing to their faint γ -ray and afterglow emission, which challenge our current detection limits. The Swift satellite mission, with its higher γ -ray sensitivity (compared to current missions) and improved localization capability (enabling rapid identification of afterglow counterparts), is expected to revolutionize our understanding of cosmic explosions. $\square$

Received 25 March; accepted 15 June 2004; doi:10.1038/nature02757.

- Pain, E. *et al.* BEPOSAX observations of GRB 980425: Detection of the prompt event and monitoring of the error box. *Astrophys. J.* **536**, 778–787 (2000).
- Galama, T. J. *et al.* An unusual supernova in the error box of the gamma-ray burst of 25 April 1998. *Nature* **395**, 670–672 (1998).
- Kulkarni, S. R. *et al.* Radio emission from the unusual supernova 1998bw and its association with the gamma-ray burst of 25 April 1998. *Nature* **395**, 663–669 (1998).
- Li, Z. & Chevalier, R. A. Radio supernova SN 1998BW and its relation to GRB 980425. *Astrophys. J.* **526**, 716–726 (1999).
- Gotz, D., Mereghetti, S., Beck, M., Borkowski, J. & Mowlavi, N. GRB 031203: a long GRB detected with INTEGRAL GRB Circ. Netw. 2459 (2003).
- Prochaska, J. X. *et al.* The host galaxy of GRB 031203: Implications of its low metallicity, low redshift, and starburst nature. *Astrophys. J.* (submitted).
- Sazonov, S. Y., Lutoninov, A. A. & Sunyaev, R. A. An apparently normal gamma-ray burst with an unusually low luminosity. *Nature* doi:10.1038/nature02748 (this issue).
- Bloom, J. S. *et al.* Expected characteristics of the subclass of supernova gamma-ray bursts. *Astrophys. J.* **506**, L105–L108 (1998).
- Norris, J. P. Implications of the lag-luminosity relationship for unified gamma-ray burst paradigms. *Astrophys. J.* **579**, 386–403 (2002).
- Santos-Lleo, M., Calderon, P. & Gotz, D. GRB 031203 XMM-Newton observation GRB Circ. Netw. 2464 (2003).
- Watson, D. *et al.* A very low luminosity X-ray flash: XMM-Newton observations of GRB 031203. *Astrophys. J. Lett.* **605**, L101–L104 (2004).
- Rodriguez-Pascual, P., Santos-Lleo, M., Gonzalez-Riestra, R., Schartel, N. & Altieri, B. XMM-Newton observation of GRB031203 GRB Circ. Netw. 2474 (2003).

- Gal-Yam, A. The near infra-red light curve of SN 2003lw, associated with GRB 031203. *Astrophys. J.* **609**, L59–L62 (2004).
- Vanderspek, R. *et al.* Hete observations of the gamma-ray burst GRB030329: Evidence for an underlying soft x-ray component. Preprint astro-ph/0401311 at (<http://xxx.lanl.gov>) (2004).
- Berger, E., Kulkarni, S. R. & Frail, D. A. A standard kinetic energy reservoir in gamma-ray burst afterglows. *Astrophys. J.* **590**, 379–385 (2003).
- Frail, D. A., Kulkarni, S. R., Berger, E. & Wieringa, M. H. A complete catalog of radio afterglows: The first five years. *Astron. J.* **125**, 2299–2306 (2003).
- Panaitecu, A. & Kumar, P. Properties of relativistic jets in gamma-ray burst afterglows. *Astrophys. J.* **571**, 779–789 (2002).
- Sari, R., Piran, T. & Narayan, R. Spectra and light curves of gamma-ray burst afterglows. *Astrophys. J.* **497**, L17–L20 (1998).
- Granot, J. & Sari, R. The shape of spectral breaks in gamma-ray burst afterglows. *Astrophys. J.* **586**, 820–829 (2002).
- Frail, D. A. *et al.* Beaming in gamma-ray bursts: Evidence for a standard energy reservoir. *Astrophys. J.* **562**, L55–L58 (2001).
- Waxman, E. The nature of GRB980425 and the search for off-axis GRB signatures in nearby type Ib/c supernovae emission. *Astrophys. J.* **602**, 886–891 (2004).
- Frail, D. A., Waxman, E. & Kulkarni, S. R. A 450 day light curve of the radio afterglow of GRB 970508: Fireball calorimetry. *Astrophys. J.* **537**, 191–204 (2000).
- Amati, L. *et al.* Intrinsic spectra and energetics of BeppoSAX gamma-ray bursts with known redshifts **390**, 81–89 (2002).
- Nakamura, T., Mazzali, P. A., Nomoto, K. & Iwamoto, K. Light curve and spectral models for the hypernova SN 1998BW associated with GRB 980425. *Astrophys. J.* **550**, 991–999 (2001).
- Soderberg, A. M., Frail, D. A. & Wieringa, M. A. Severe constraints on off-axis GRB jets in type Ib/c supernovae from late-time radio observations. *Astrophys. J.* **607**, L13–L16 (2004).
- Bloom, J. S., Frail, D. A. & Kulkarni, S. R. Gamma-ray burst energetics and the gamma-ray burst Hubble diagram: Promises and limitations. *Astrophys. J.* **594**, 674–683 (2003).
- Berger, E. *et al.* A common origin for cosmic explosions inferred from calorimetry of GRB030329. *Nature* **426**, 154–157 (2003).
- Soderberg, A. M. *et al.* A redshift determination for XRF 020903: First spectroscopic observations of an X-ray flash. *Astrophys. J.* **606**, 994–999 (2004).
- Berger, E., Kulkarni, S. R., Frail, D. A. & Soderberg, A. M. A radio survey of type Ib and Ic supernovae: Searching for engine-driven supernovae. *Astrophys. J.* **599**, 408–418 (2003).

Acknowledgements GRB research at Caltech is supported in part by NSF and NASA. We are indebted to S. Barthelmy and the GCN. The VLA is operated by the National Radio Astronomy Observatory, a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc. A.M.S. is supported by an NSF graduate research fellowship. A.G. acknowledges support by NASA through a Hubble fellowship grant.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.M.S. (ams@astro.caltech.edu).

Two-dimensional geometry of spin excitations in the high-transition-temperature superconductor $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$

V. Hinkov¹, S. Pailhès², P. Bourges², Y. Sidis², A. Ivanov³, A. Kulakov¹, C. T. Lin¹, D. P. Chen¹, C. Bernhard¹ & B. Keimer¹

¹Max-Planck-Institut für Festkörperforschung, 70569 Stuttgart, Germany

²Laboratoire Léon Brillouin, CEA-CNRS, CE-Saclay, 91191 Gif-sur-Yvette, France

³Institut Laue-Langevin, 156X, 38042 Grenoble cedex 9, France

The fundamental building block of the copper oxide superconductors is a Cu_4O_4 square plaquette. The plaquettes in most of these materials are slightly distorted to form a rectangular lattice, for which an influential theory predicts that high-transition-temperature (high- T_c) superconductivity is nucleated in 'stripes' aligned along one of the axes^{1–3}. This theory received strong support from experiments that indicated a one-dimensional character for the magnetic excitations in the high- T_c material $\text{YBa}_2\text{Cu}_3\text{O}_{6.6}$ (ref. 4). Here we report neutron scattering data on 'untwinned' $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ crystals, in which the orientation of the rectangular lattice is maintained throughout the entire

volume. Contrary to the earlier claim⁴, we demonstrate that the geometry of the magnetic fluctuations is two-dimensional. Rigid stripe arrays therefore appear to be ruled out over a wide range of doping levels in $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ but the data may be consistent with liquid-crystalline stripe order⁵. The debate about stripes has therefore been reopened.

Inelastic neutron scattering data can pinpoint the theoretically predicted stripe phases by virtue of their characteristic quasi-one-dimensional spin excitations⁶. Spin excitations incommensurate with the crystal lattice have indeed been observed by neutron scattering in several families of layered copper oxides^{4,7–12}. However, almost all neutron experiments thus far reported have been carried out on fully ‘twinned’ crystals with equal proportions of micrometre-scale twin domains in which the rectangular Cu_4O_4 plaquettes are rotated by 90° with respect to one another^{7–11}. Because the scattering pattern from such crystals consists of equal contributions from both twin domains, even perfectly one-dimensional spin fluctuations generate a fourfold symmetric pattern, so that they cannot be discriminated from microscopically two-dimensional fluctuations. The results of two neutron scattering experiments on partially detwinned $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ crystals^{4,12} have been interpreted as evidence of a one-dimensional character of the magnetic fluctuations^{4,12–14}. However, owing to significant contributions from the minority domain, the full geometry of the excitation spectrum has remained unclear.

Using neutron scattering from a mosaic of fully untwinned, nearly optimally doped $\text{YBa}_2\text{Cu}_3\text{O}_{6.85}$ crystals (see Methods) we have resolved this issue. We find that the locus of maximum spin fluctuation spectral weight approximately forms a circle in momentum space, so that the basic character of the spin excitations is two-dimensional. However, the damping and amplitude along the circle are modulated in a one-dimensional fashion. The strength of this modulation depends strongly on the excitation energy. Data from underdoped, untwinned $\text{YBa}_2\text{Cu}_3\text{O}_{6.6}$ crystals exhibit similar features.

Scans through the (200) and (020) Bragg reflections of the $\text{YBa}_2\text{Cu}_3\text{O}_{6.85}$ array reveal a bulk population ratio between majority

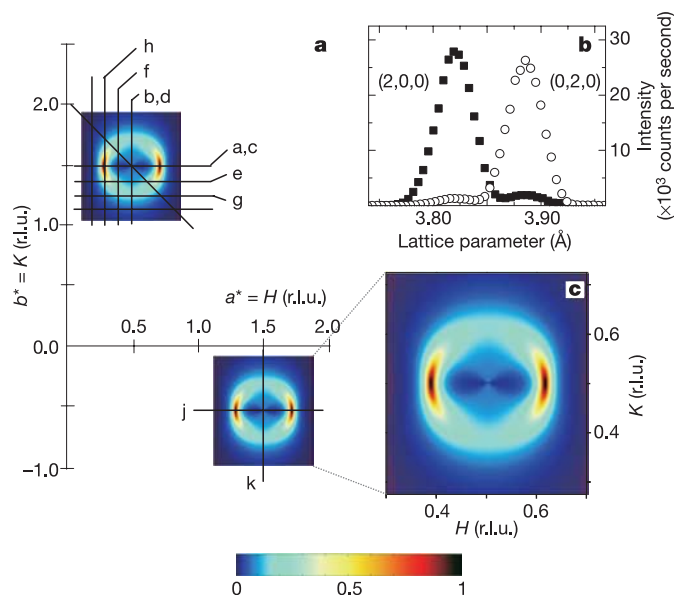


Figure 1 Layout of the reciprocal lattice and magnetic spectral weight of $\text{YBa}_2\text{Cu}_3\text{O}_{6.85}$. **a**, In-plane projection of the $\text{YBa}_2\text{Cu}_3\text{O}_{6.85}$ reciprocal lattice indicating the trajectories of the constant-energy scans shown in Fig. 2. The representation is not true to scale. **b**, Longitudinal elastic scans through the (2, 0, 0) and (0, 2, 0) crystallographic Bragg reflections, demonstrating a twin domain population ratio of $\sim 95:5$. **c**, Intrinsic magnetic spectral weight at 35 meV extracted from the analysis presented in the Supplementary Information.

and minority twin domains of about 95:5 (Fig. 1b and Methods). This is one order of magnitude larger than in previous experiments^{4,12}. The contribution of the minority domain to the magnetic scattering pattern is thus negligible in our experiments.

Figures 2 and 3 show magnetic neutron scattering data from this crystal array. The overall features of the neutron cross-section are in good agreement with prior work on twinned crystals^{7,9–11}. Owing to antiferromagnetic (AF) correlations between spins in the CuO_2 plane, the spectral weight of low-energy spin excitations in $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ is centred around the in-plane wave vector $\mathbf{Q}_{\text{AF}} = (0.5, 1.5)$. As a consequence of antiferromagnetic interactions within a bilayer unit, the magnetic intensity is modulated as a function of the c -axis wave vector transfer, L , such that the maximum intensity is observed at odd multiples of $L = 1.7$. The characteristic L -dependence of the intensity implies that the signal originates from the CuO_2 planes. As observed before, the magnetic excitations are strongly enhanced below the superconducting transition temperature. For excitation energies below the ‘spin gap’, $\hbar\omega_g = 30$ meV, the magnetic spectral weight is below the detection limit. The lowest-energy excitations observed above $\hbar\omega_g$ are incommensurate (Fig. 2). The incommensurate excitation branches disperse towards $\mathbf{Q}_{\text{AF}}$ with increasing excitation energy (Fig. 3), and they merge at $\hbar\omega_0 = 41$ meV, giving rise to the ‘resonance peak’.

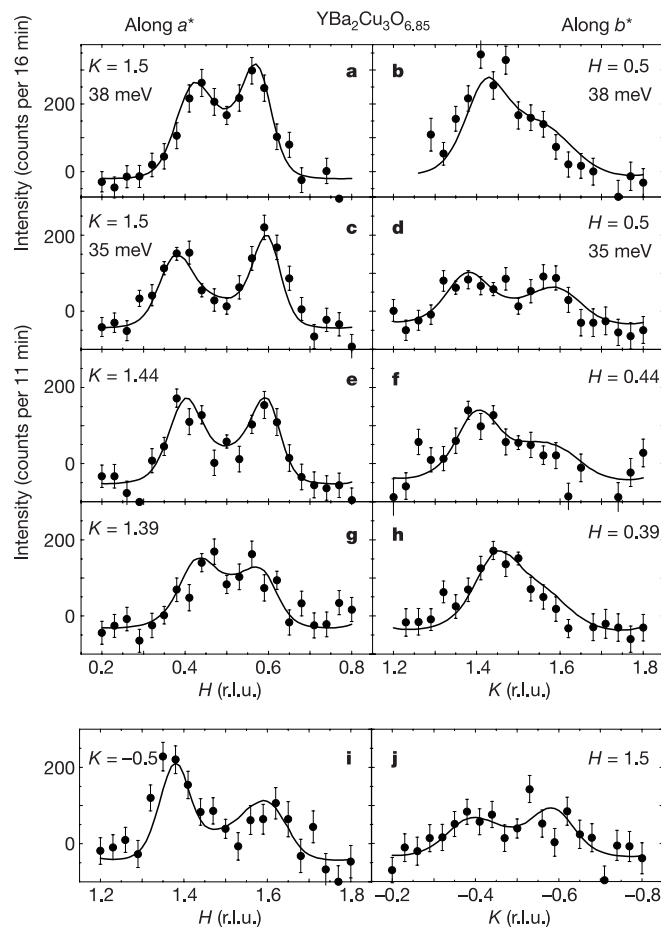


Figure 2 Constant-energy scans along the trajectories indicated in Fig. 1a. The excitation energies are $\hbar\omega = 38$ meV (**a**, **b**) and 35 meV (**c**–**j**). The wave vector $\mathbf{Q} = (H, K, 1.7)$ is given in reciprocal lattice units, r.l.u. (see Methods). We show subtractions of the intensities at $T = 10$ K ($< T_c$) and $T = 100$ K ($> T_c$). The lines result from a convolution of a model cross-section with the instrumental resolution function (see Methods). The data in panels **c**, **d** and **i**, **j** were taken in two different Brillouin zones with exchanged resolution conditions. The observed anisotropy between a^* and b^* is thus not due to resolution effects. The error bars represent the square root of the number of detector counts.

The new aspect of this work is the determination of the in-plane geometry of the spin excitations. Figure 2c–j shows representative scans from a comprehensive map of the spin fluctuation spectral weight at $\hbar\omega = 35$ meV. More limited data sets were also taken at $\hbar\omega = 31, 33$ and 37 (not shown), and 38 meV (Fig. 2a, b). To extract the magnetic spectral weight from the experimentally determined scattering profiles, we have numerically convoluted an anisotropic damped harmonic oscillator cross-section with the spectrometer resolution function (see Supplementary Information). The computed profiles, shown as solid lines in Figs 2 and 3, provide excellent descriptions of the experimental data. The intrinsic magnetic spectral weight at 35 meV extracted from this analysis is depicted in Fig. 1c.

We emphasize two salient features of the magnetic spectrum of $\text{YBa}_2\text{Cu}_3\text{O}_{6.85}$. The most important observation is that well-defined incommensurate peaks are present in scans along both a^* and b^* (Fig. 2a–d). This demonstrates the intrinsic two-dimensional nature of the spin excitations. The displacements from the commensurate wave vector $\mathbf{Q}_{\text{AF}}$ along a^* and b^* are identical within an experimental error of 15% at all energies covered by our experiment. The locus of maximum magnetic intensity approximately forms a circle around $\mathbf{Q}_{\text{AF}}$ following an isotropic dispersion relation $\omega_{\mathbf{Q}} = \omega_0 - c(\mathbf{Q} - \mathbf{Q}_{\text{AF}})^2$ with parameters $\hbar\omega_0 = 41 \pm 1$ meV and $\hbar c = (180 \pm 30)$ meV Å² for all energies.

Second, both the amplitude and the width of the incommensurate peaks exhibit pronounced in-plane anisotropies (Fig. 1c), which are strongly energy-dependent. At $\hbar\omega = 35$ meV, for instance, the scans along a^* (Fig. 2c) are resolution-limited, but those along b^* (Fig. 2d) are much broader. However, the $|\mathbf{Q}|$ -integrated spectral weights along a^* and b^* differ by at most 40%. At $\hbar\omega = 38$ meV, the width anisotropy is significantly smaller than at 35 meV (Fig. 2a, b), and the $|\mathbf{Q}|$ -integrated intensities along a^* and b^* are identical to within 15%.

To check the generality of our findings, we have repeated our experiment on an array of $\text{YBa}_2\text{Cu}_3\text{O}_{6.6}$ crystals ($T_c = 61$ K) with a 94:6 detwinning ratio. Representative data are shown in Fig. 4. In accord with prior work on twinned crystals^{7,9,12}, the resonance peak is observed at a lower energy, $\hbar\omega_0 = 37.5$ meV, than at optimum

doping. The salient features of the incommensurate excitations below $\hbar\omega_0$ are, however, quite similar. Specifically, incommensurate peaks are observed along both a^* and b^* at all energies monitored in our experiment, so that the geometry of the excitation spectrum is two-dimensional. The amplitude and width anisotropy is comparable to that at optimum doping, and it exhibits the same increase with decreasing energy. Even at 28 meV, that is, 10 meV below the resonance peak, the $|\mathbf{Q}|$ -integrated spectral weights along a^* and b^* differ by at most 50%. The only significant difference is that the dispersion relation is somewhat steeper along b^* than along a^* , whereas it is isotropic at optimum doping.

An anisotropy of the magnetic signal along a^* and b^* was first pointed out by Mook *et al.*⁴ for a $\text{YBa}_2\text{Cu}_3\text{O}_{6.6}$ crystal with a detwinning ratio of 2:1. From a limited data set at a single energy, they deduced a 100% intensity difference between the two directions and interpreted their data as evidence of a one-dimensional spin fluctuation geometry. Using an analysis of a large set of data between $\hbar\omega_g$ and $\hbar\omega_0$, we have now, however, demonstrated that the low-energy spin excitations of both underdoped and optimally doped $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ have a two-dimensional geometry. Which models can describe the observed one-dimensional amplitude and width anisotropy?

Theories based on a one-dimensional, rigid array of stripes predict a 100% intensity anisotropy and cannot account for the two-dimensional scattering pattern^{13,14}. The map of the magnetic intensity at 35 meV does, however, bear a resemblance to the scattering pattern generated by a nematic liquid crystal close to a nematic-to-smectic critical point. In this scenario, the structural anisotropy between the a and b axes may act as an aligning field for the nematic director⁶. A preferential alignment of charge stripes along b would enhance the magnetic spectral weight along a . However, detailed theoretical work is required to ascertain whether the observed scattering pattern is indeed a unique signature of a quantum fluid of fluctuating stripes. A theoretical description in this framework should also account for the observed energy dependence of the anisotropy. An alternative approach would be to consider a two-dimensional arrangement of charged domain walls^{1,15}.

In the absence of experimental information about the in-plane

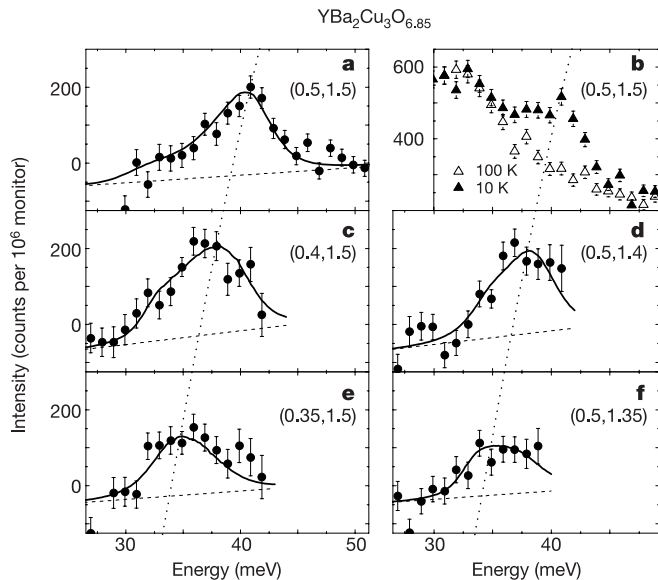


Figure 3 Constant- $\mathbf{Q}$ scans demonstrating the dispersion of the magnetic mode along a^* (a, c, e) and b^* (b, d, f). Panel b shows raw data above and below T_c . The remaining panels show subtractions of the intensities at $T = 10$ K ($< T_c$) and $T = 100$ K ($> T_c$). The solid lines result from a numerical convolution of a model cross-section with the instrumental resolution function (see Methods). The dashed lines show the slightly energy-dependent background due to the thermal population of phonon excitations. The error bars represent the square root of the number of detector counts.

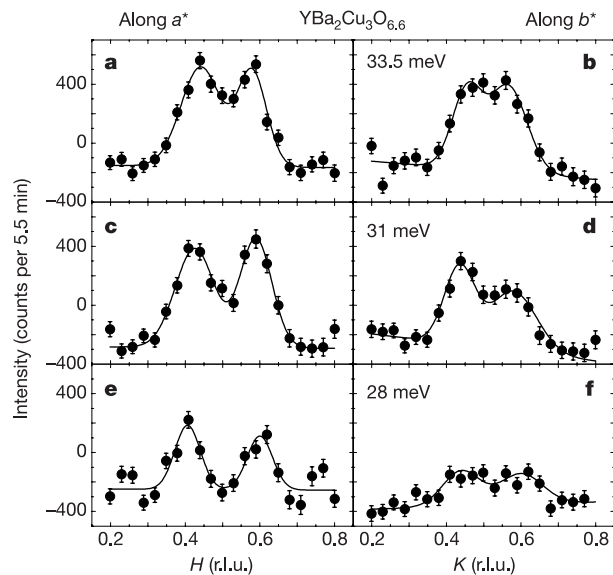


Figure 4 Constant-energy scans for $\text{YBa}_2\text{Cu}_3\text{O}_{6.6}$. They were performed along a^* at $\mathbf{Q} = (H, -1.5, 1.7)$ (a, c, e) and along b^* at $\mathbf{Q} = (1.5, K, 1.7)$ (b, d, f), at different energies $\hbar\omega$. The plots show subtractions of the intensities at $T = 5$ K and $T = 250$ K. The lines are double gaussians fitted to the data. The resolution conditions were identical for measurements along a^* and b^* , so that the scans are directly comparable. The error bars represent the square root of the number of detector counts.

anisotropy, prior Fermi-liquid-based theoretical scenarios for the spin dynamics of $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ had only considered the influence of the CuO_2 layers, ignoring a possible influence of the b -axis-oriented CuO chains^{16,17}. On the basis of the observed c -axis modulation of the intensity, we can rule out a direct contribution from the CuO chains to the observed magnetic signal. However, an indirect influence of the CuO chains on the spin dynamics in the layers cannot be excluded. The plane-derived Fermi surfaces are predicted to exhibit a small in-plane anisotropy owing to the hybridization between orbitals centred on chains and layers and the different in-plane Cu-O bond lengths along a and b ¹⁸. Further, nuclear quadrupole resonance experiments¹⁹ show that charge-density waves in the CuO chains may induce a charge modulation in the CuO_2 layers, and X-ray scattering experiments²⁰ indicate a polarization of Cu orbitals in the CuO_2 layers that is due to oxygen ordering in the chain layers. The impact of these factors on the spin dynamics will have to be assessed in quantitative calculations. Other factors, such as proximity to a 'Pomeranchuk' instability of the Fermi surface^{21–23}, may also contribute to the anisotropy of the spin dynamics. The observed in-plane anisotropy of the superconducting energy gap imposes further constraints on the theory^{24,25}.

In conclusion, our data put stringent, quantitative constraints on stripe theories of high- T_c superconductivity. They rule out a large class of theories according to which high- T_c superconductivity coexists with a rigid array of stripes, over a wide range of hole concentrations encompassing the 60 and 90 K superconducting phases of $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$. If interpreted as evidence of the theoretically predicted liquid-crystalline stripe phases, they quantify the orientational stripe fluctuations. □

Methods

Sample preparation

High-quality, single-phase $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ crystals were synthesized by the solution growth method and cut into rectangular shapes. For the nearly optimally doped sample, 82 crystals with a total mass of ~ 1.3 g were annealed at 520 °C in O_2 , resulting in an oxygen content of $x \approx 0.85$. The crystals were individually detwinned by subjecting them to a uniaxial mechanical stress of $\sim 5 \times 10^7 \text{ N m}^{-2}$ along the (100) axis at 520 °C in flowing oxygen for at least 4 h (refs 26, 27). The resulting specimens show superconducting transition temperatures, T_c , of 90 K with ΔT_c between 1 and 3 K. They were then co-aligned on two Al plates into which a dense grid of holes had been drilled. By using thin Al wire to fix the crystals to the grid, it proved possible largely to avoid organic glue, and hence substantially to reduce the background signal due to incoherent neutron scattering from hydrogen. The full array shows a mosaicity of $< 1.2^\circ$ and is almost entirely untwinned (Fig. 1b). The underdoped sample consists of 120 crystals with a total mass of ~ 2 g. After annealing at 590 °C in air and detwining at 400 °C in Ar, these specimens show a T_c of 61 K with ΔT_c between 1 and 2 K. They were co-aligned on two Al plates and fixed by Al screws, thus completely avoiding organic glue. The mosaicity is $\sim 1.2^\circ$ and the domain population ratio 94:6.

Neutron scattering

The wave vector $\mathbf{Q} = (H, K, L)$ is quoted in units of the reciprocal lattice vectors a^* , b^* and c^* , where $a = 2\pi/a^* = 3.82 \text{ \AA}$, $b = 2\pi/b^* = 3.88 \text{ \AA}$, and $c = 2\pi/c^* = 11.7 \text{ \AA}$. Because of the slightly different lengths of the lattice parameters a and b in the CuO_2 planes, the (200) and (020) crystallographic Bragg reflections occur at different wave vector transfers $\mathbf{Q}$. Because neutron scattering is a bulk probe, scans through these reflections directly reveal the bulk population ratio of the two twin domains, as shown in Fig. 1b.

The data for nearly optimally doped $\text{YBa}_2\text{Cu}_3\text{O}_{6.85}$ were collected on the 2T triple-axis spectrometer at the Laboratoire Léon Brillouin (LLB), Saclay, France. Because of kinematical restrictions, the scans were carried out in the second Brillouin zone (Fig. 1a). A specially designed holder was used, which enabled us to access the $P_+ = \{(1, 0, 0); (0, 3, 3.4)\}$ scattering plane as well as the complementary $P_- = \{(0, -1, 0); (3, 0, 3.4)\}$ plane during the experiment without opening the cooling chamber. This change between P_+ and P_- is equivalent to a rotation of the resolution ellipsoid by 90° with respect to a^* and b^* and allows a quick check for resolution effects. No collimators were used, in order to maximize the neutron flux. Graphite filters extinguished higher-order contamination of the neutron beam. The data for underdoped $\text{YBa}_2\text{Cu}_3\text{O}_{6.6}$ were collected on the IN8 spectrometer at the Institut Laue-Langevin, Grenoble, France. The same holder was employed and the spectrometer was configured in the same way. The final wave vector was fixed to either $2.662 \text{ \AA}^{-1}$ (Figs 2, 3c–f, 4) or $3.85 \text{ \AA}^{-1}$ (Fig. 3a, b).

Data analysis

Because the magnetic excitations are overdamped at high temperatures, the spin fluctuation contribution to the neutron cross-section is most clearly apparent in temperature subtractions. For $\text{YBa}_2\text{Cu}_3\text{O}_{6.85}$ we show subtractions between the intensities

at $T = 10$ and 100 K. For $\text{YBa}_2\text{Cu}_3\text{O}_{6.6}$, which exhibits a significant magnetic signal in the normal state above $T_c = 61$ K, we show subtractions between $T = 5$ and 250 K. The simplest model consistent with our data consists of an anisotropic damped harmonic oscillator with an isotropic dispersion law, isotropic distribution of the $|\mathbf{Q}|$ -integrated intensity, and anisotropic damping parameter in the a^*-b^* plane (see Supplementary Information). The computed profiles shown in Figs 2 and 3 are numerical convolutions of this model with the spectrometer resolution function. The numerical resolution convolution was performed with the program Rescal-5 based on the Popovici method²⁸ (see online description under <http://www.ill.fr/tas/matlab/doc/rescal5/rescal.htm>) as well as with in-house programs at LLB.

Received 18 March; accepted 22 June 2004; doi:10.1038/nature02774.

- Zaenen, J. & Gunnarson, O. Charged magnetic domain lines and the magnetism of high- T_c oxides. *Phys. Rev. B* **40**, 7391–7394 (1989).
- Emery, V. J. & Kivelson, S. A. Frustrated electronic phase separation and high-temperature superconductors. *Physica C* **209**, 597–621 (1993).
- Tranquada, J. M., Sternlieb, B. J., Axe, J. D., Nakamura, Y. & Uchida, S. Evidence for stripe correlations of spins and holes in copper oxide superconductors. *Nature* **375**, 561–563 (1995).
- Mook, H. A., Dai, P., Dogan, F. & Hunt, R. D. One-dimensional nature of the magnetic fluctuations in $\text{YBa}_2\text{Cu}_3\text{O}_{6.6}$. *Nature* **404**, 729–731 (2000).
- Kivelson, S. A., Fradkin, E. & Emery, V. J. Electronic liquid-crystal phases of a doped Mott insulator. *Nature* **393**, 550–553 (1998).
- Kivelson, S. A. *et al.* How to detect fluctuating stripes in the high-temperature superconductors. *Rev. Mod. Phys.* **75**, 1201–1241 (2003).
- Mook, H. A. *et al.* Spin fluctuations in $\text{YBa}_2\text{Cu}_3\text{O}_{6.6}$. *Nature* **395**, 580–582 (1998).
- Yamada, K. *et al.* Doping dependence of the spatially modulated dynamical spin correlations and the superconducting transition temperature in $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$. *Phys. Rev. B* **57**, 6165–6172 (1998).
- Arai, M. *et al.* Incommensurate spin dynamics of underdoped superconductor $\text{YBa}_2\text{Cu}_3\text{O}_{6.7}$. *Phys. Rev. Lett.* **83**, 608–611 (1999).
- Bourges, P. *et al.* The spin excitation spectrum in superconducting $\text{YBa}_2\text{Cu}_3\text{O}_{6.85}$. *Science* **288**, 1234–1237 (2000).
- Reznik, D. *et al.* Dispersion of magnetic excitations in superconducting optimally doped $\text{YBa}_2\text{Cu}_3\text{O}_{6.95}$. Preprint at (<http://xxx.arxiv.org/pdf/cond-mat/0307591>) (2003).
- Stock, C. *et al.* Dynamic stripes and resonance in the superconducting and normal phases of $\text{YBa}_2\text{Cu}_3\text{O}_{6.5}$ ortho-II superconductor. *Phys. Rev. B* **69**, 014502 (2004).
- Batista, C. D., Ortiz, G. & Balatsky, A. V. Unified description of the resonance peak and incommensuration in high- T_c superconductors. *Phys. Rev. B* **64**, 172508 (2001).
- Krüger, F. & Scheidl, S. Spin dynamics of stripes. *Phys. Rev. B* **67**, 134512 (2003).
- Fine, B. V. Hypothesis of two-dimensional stripe arrangement and its implications for the superconductivity in high- T_c cuprates. Preprint at (<http://xxx.arxiv.org/pdf/cond-mat/0308428>) (2003).
- Brinckmann, J. & Lee, P. A. Renormalized mean-field theory of neutron scattering in cuprate superconductors. *Phys. Rev. B* **65**, 014502 (2001).
- Onufrieva, F. & Pfeuty, P. Spin dynamics of a two-dimensional metal in a superconducting state: Application to the high- T_c cuprates. *Phys. Rev. B* **65**, 054515 (2002).
- Andersen, O. K., Liechtenstein, A. I., Jepsen, O. & Paulsen, F. LDA energy bands, low-energy Hamiltonians, t' , t'' , $t_{\text{perp}}(k)$ and $J_{\text{perp}}(k)$. *J. Phys. Chem. Solids* **56**, 1573–1591 (1995).
- Grévin, B., Berthier, Y. & Collon, G. In-plane charge modulation below T_c and charge density-wave correlations in the chain layer in $\text{YBa}_2\text{Cu}_3\text{O}_7$. *Phys. Rev. Lett.* **85**, 1310–1313 (2000).
- Feng, D. L. *et al.* Zooming-in on the charge ordering in $\text{YBa}_2\text{Cu}_3\text{O}_{6.5}$. Preprint at (<http://xxx.arxiv.org/pdf/cond-mat/0402488>) (2004).
- Yamase, H. & Kohno, H. Possible quasi-one-dimensional Fermi surface in $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$. *J. Phys. Soc. Jpn* **69**, 332–335 (2000).
- Halboth, C. J. & Metzner, W. D-wave superconductivity and Pomeranchuk instability in the two-dimensional Hubbard model. *Phys. Rev. Lett.* **85**, 5162–5165 (2000).
- Oganesyan, V., Kivelson, S. A. & Fradkin, E. Quantum theory of a nematic Fermi liquid. *Phys. Rev. B* **64**, 195109 (2001).
- Limonov, M. F., Rykov, A. I., Tajima, S. & Yamanaka, A. Superconductivity-induced effects on phononic and electronic Raman scattering in twin-free $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ single crystals. *Phys. Rev. B* **61**, 12412–12419 (2000).
- Lu, D. H. *et al.* Superconducting gap and strong in-plane anisotropy in untwinned $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$. *Phys. Rev. Lett.* **86**, 4370–4373 (2001).
- Lin, C. T., Zhou, W. & Liang, W. Y. Growth of large and untwinned single crystals of YBCO. *Physica C* **195**, 291–300 (1992).
- Voronkova, V. I. & Wolf, T. Thermomechanical detwining of $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ single crystals under reduced oxygen partial pressure. *Physica C* **218**, 175–180 (1993).
- Popovici, M. On the resolution of slow-neutron spectrometers. IV. The triple-axis spectrometer resolution function, spatial effects included. *Acta Crystallogr. A* **31**, 507–513 (1975).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank B. Hennion, S. Kivelson, D. Manske, W. Metzner and H. Yamase for discussions, S. Lacher, H. Wendel, B. Baum and M. Bakr for crystal preparation, M. Ohl and W. Plener for the design and manufacturing of the sample holder, H. Bender, C. Busch and H. Klann for technical support, and P. Baroni for technical support at the 2T spectrometer. We acknowledge support from the German Federal Ministry of Culture and Research (BMBF) and from the Deutsche Forschungsgemeinschaft.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to B.K. (b.keimer@fkf.mpg.de).

Controlling the dynamics of spontaneous emission from quantum dots by photonic crystals

Peter Lodahl¹, A. Floris van Driel², Ivan S. Nikolaev¹, Arie Iman¹, Karin Overgaag¹, Daniël Vanmaekelbergh² & Willem L. Vos¹

¹Complex Photonic Systems (COPS), Department of Science and Technology and MESA⁺ Research Institute, University of Twente, PO Box 217, 7500 AE Enschede, The Netherlands

²Debye Institute, Condensed Matter and Interfaces, Utrecht University, PO Box 80.000, 3508 TA Utrecht, The Netherlands

Control of spontaneously emitted light lies at the heart of quantum optics. It is essential for diverse applications ranging from miniature lasers and light-emitting diodes^{1–5}, to single-photon sources for quantum information^{6–8}, and to solar energy harvesting⁹. To explore such new quantum optics applications, a suitably tailored dielectric environment is required in which the vacuum fluctuations that control spontaneous emission can be manipulated^{10,11}. Photonic crystals provide such an environment: they strongly modify the vacuum fluctuations, causing the decay of emitted light to be accelerated or slowed down^{12,13}, to reveal unusual statistics¹⁴, or to be completely inhibited in the ideal case of a photonic bandgap^{1,15}. Here we study spontaneous emission from semiconductor quantum dots embedded in inverse opal photonic crystals¹⁶. We show that the spectral distribution and time-dependent decay of light emitted from excitons confined in the quantum dots are controlled by the host photonic crystal. Modified emission is observed over large frequency bandwidths of 10%, orders of magnitude larger than reported for resonant optical microcavities¹⁷. Both inhibited and enhanced decay rates are observed depending on the optical emission frequency, and they are controlled by the crystals' lattice parameter. Our experimental results provide a basis for all-solid-state dynamic control of optical quantum systems¹⁸.

The characteristics of spontaneously emitted light depend strongly on the environment of the emitter; both the total rate of emission and the direction-dependent spectra are affected. In the Wigner–Weisskopf approximation¹⁰, the radiative decay rate is described by Fermi's 'golden rule'¹¹ that contains the local density of electromagnetic states^{19,20} (LDOS). The LDOS counts the available number of electromagnetic modes in which photons can be emitted at the specific location of the emitter, and can be interpreted as the density of vacuum fluctuations. The LDOS and hence the emission rate depend sensitively on the emission frequency ω and on the position of the emitter, but not on the emission direction. Here we report on the frequency-dependent emission rate of quantum dots under the influence of modified vacuum fluctuations in photonic crystals.

In photonic crystals (Fig. 1a), emission of light is modified because the refractive index varies spatially on length scales of the order of the wavelength. Hence, light interference by Bragg diffraction causes characteristic stopbands leading to direction-dependent spectra. The emission rate is modified if Bragg stop bands exclude a substantial solid angle of the crystal. To confirm the influence of Bragg diffraction, we have performed angle-resolved measurements of the emission spectra of light sources embedded in titania inverse opals (see Methods). As light sources, we use colloidal CdSe quantum dots that are highly efficient, narrow linewidth, size-dependent emitters that effectively resemble two-level systems. Figure 1b shows the emission spectra for a sample with lattice parameter $a = 420$ nm for different detection angles α relative to the (111) planes. An unusual angular dependence is observed

because the emitted intensity is strongly attenuated at $\alpha = 0^\circ$, increases at $\alpha = 20^\circ$ and $\alpha = 50^\circ$, and decreases again at $\alpha = 60^\circ$. This behaviour is caused by a photonic stopband centred at $\alpha = 0^\circ$ that is instantly recognized when the emission spectra are referenced to spectra measured at $\alpha = 60^\circ$. The latter are suitable references because here the stopbands are shifted far outside the emission spectrum of the quantum dots.

Figure 1c displays the angular dependence of the relative intensity for samples with three different lattice parameters. For a reference

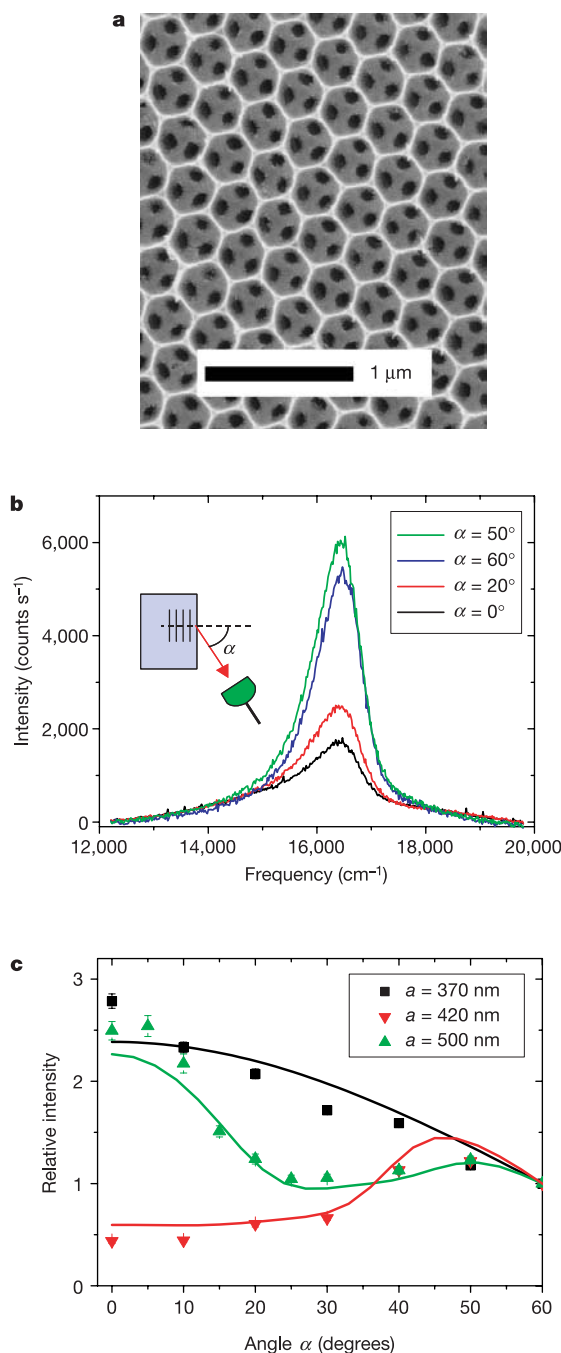


Figure 1 Angle-resolved spectral measurements. **a**, Scanning electron microscope image of the (111) face of a titania inverse opal with lattice parameter $a = 460$ nm. **b**, Recorded quantum dot emission spectra from an inverse opal with lattice parameter $a = 420$ nm for different detection angles α relative to the (111) lattice planes. **c**, Measured (points) and calculated (curves) angle-dependent intensities at fixed frequency $\omega = 15,870$ cm⁻¹ for three samples with different lattice parameters. The intensities are divided by the intensity measured at $\alpha = 60^\circ$, where no stopbands appear.

sample with a small lattice parameter ($a = 370$ nm), the stopgaps appear at higher frequencies than the emission spectrum of the quantum dots and the intensity decays monotonically with increasing angle. In contrast, the $a = 420$ nm and $a = 500$ nm samples show pronounced photonic effects with strongly reduced emission in a broad stopband centred at $\alpha = 0^\circ$ and $\alpha = 25^\circ$, respectively. We compared the measured angle-dependent intensities to a recently developed theory that describes propagation of light that is diffused by ubiquitous disorder present in real photonic crystals²¹. The curves in Fig. 1c are the predictions with no adjustable parameters, using only the measured reflectivity peaks and calculated band-structure as input. We obtain excellent agreement between experiment and theory, confirming that the emission of the embedded quantum dots is strongly affected by the photonic crystal.

The decay rate of the excited state of the quantum dots in the photonic crystal was measured by exciting with a short optical pulse and recording the spontaneous emission as a function of time. In photonic crystals, the LDOS and (as a result) the decay rate ideally vanish over a frequency range known as the photonic bandgap. In real crystals with finite dimensions, the LDOS is never zero, but strongly modulated pseudogaps are anticipated^{19,20}. In such gaps, vacuum fluctuations are expelled from the crystal, and the spontaneous decay of an excited emitter is inhibited. Away from a gap, the LDOS is increased and spontaneous emission will be enhanced. Figure 2 shows decay curves measured on three different photonic crystals at selected optical frequencies. Two main contributions are present in the data: a fast decay (~ 1 ns) and a slow decay (>10 ns). The fast decay is caused by defect-related luminescence of the titania backbone, as confirmed from measurements on bare titania samples. The slow decay is due to the quantum dots, and is easily distinguished from the titania emission owing to its markedly different decay time. Visual inspection of the decay curves in Fig. 2 reveals both accelerated and inhibited decay rates compared to that measured in the non-photonic sample ($a = 370$ nm). We obtain lifetimes of 9.6 ± 0.1 ns ($a = 420$ nm) and 19.3 ± 0.2 ns ($a = 500$ nm) as compared to a reference value of 12.4 ± 0.1 ns ($a = 370$ nm).

The decay curves of the quantum dots are not single exponentials, which was revealed in additional measurements over longer decay times. Non-single exponential decays in the photonic crystal might be related to fractional decay^{13,14} or a distribution of quantum dots over sites with different LDOS. However, it was recently concluded that fluctuations in the environment surrounding the quantum dots induce a broad distribution of non-radiative decay channels,

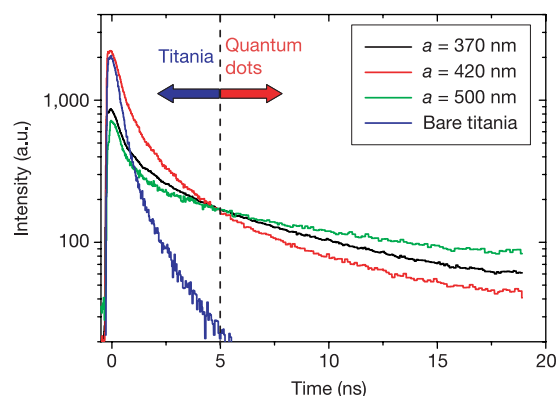


Figure 2 Luminescence decay curves of quantum dots inside three different photonic crystals. The data are recorded at frequencies $15,670$ cm^{-1} ($a = 370$ nm, black curve) and $15,100$ cm^{-1} ($a = 420$ nm, red curve, and $a = 500$ nm, green curve). The curves have been overlapped after 5 ns. The first part of the decay curve is influenced by emission of titania (blue curve, recorded at $15,400$ cm^{-1}). After 5 ns this contribution is negligible. a.u., arbitrary units.

explaining the non-single exponential decays and effectively reducing the quantum efficiency²². We find that this distribution is centred on the radiative lifetime, and has a wide asymmetric tail of lifetimes from non-radiative contributions that dominates over asymmetries induced by the photonic crystal. Consequently, we consider only photonic effects on the averaged lifetime that is obtained by investigating the quantum-dot decay curve up to ~ 20 ns, which is least sensitive to background counts and can be well approximated by a single exponential. The robustness of this method and the sensitivity to variations in the radiative decay rate were carefully checked by comparing to a fit of the decay curve based on the distribution of lifetimes. The observed lifetimes in the inverse opals are all systematically shorter than the lifetime of quantum dots in solution, which is attributed to the reduced quantum efficiency. The reduction in quantum efficiency limits the amount of enhancement and inhibition observed. Nonetheless, the data in Fig. 2 demonstrate a strong variation in the lifetime by a factor of two for photonic crystals with different lattice parameters. These measurements are the first experimental demonstration of the use of three-dimensional photonic crystals to control the radiative lifetime of emitters.

A detailed study of the frequency dependence of the total decay rate has been carried out for crystals with many different lattice parameters, four of which are shown in Fig. 3. The total decay rate is the sum of the radiative and non-radiative decay rates. For each sample, we observe a variation of the total decay rate with emission frequency of more than 50%, which indicates that the radiative decay strongly depends on the emission frequency. For comparison, the dotted orange curve in Fig. 3 is the calculated radiative decay rate for an electric dipole in a homogeneous medium, which scales with the cube of the emission frequency^{10,11}. We obtain excellent agreement with this simple frequency scaling on the non-photonic reference sample ($a = 370$ nm). A more sophisticated model for the emission of an ensemble of quantum dots would require knowledge of the size dependence of the transition-dipole moment, homogeneous broadening, and non-radiative contributions to the decay rate, and is beyond the scope of this Letter. Importantly, the non-radiative contributions are identical for each of the samples, as quantum dots from the same preparation batch were used and all photonic crystals were fabricated under identical conditions. Therefore, variations between samples with different lattice parameters are due to changes in the radiative lifetime.

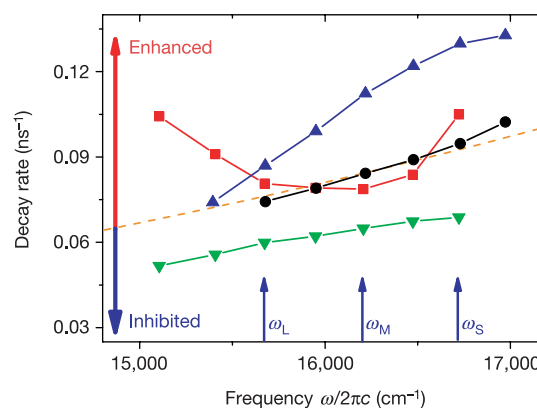


Figure 3 Measured decay rates of the excited states of quantum dots in photonic crystals with different lattice parameters. Lattice parameters are $a = 370$ nm (black circles), $a = 420$ nm (red squares), $a = 500$ nm (green upside-down triangles) and $a = 580$ nm (blue triangles). The curves connecting the measurement points are to guide the eye. The dashed curve is the calculated decay rate for a homogeneous medium. Three fixed emission frequencies are specified in the figure, corresponding to large (ω_L , diameter 6.0 nm), medium (ω_M , diameter 4.5 nm) and small (ω_S , diameter 3.8 nm) sized quantum dots.

We observe changes in the lifetime by up to a factor of two at a fixed frequency ($\sim 15,100\text{ cm}^{-1}$). For the photonic crystal with $a = 500\text{ nm}$, a pronounced inhibition of emission of 30% is observed in a wide bandwidth exceeding 10%. On the sample with $a = 420\text{ nm}$, a strongly enhanced emission is measured, extending the decay rate by more than 40% relative to the reference value. The decay rates for the $a = 580\text{ nm}$ sample are all enhanced over a broad frequency range. These measurements demonstrate that photonic crystals offer an effective way of controlling the radiative lifetime through the lattice parameter a .

To compare experimental data on different samples with theory, it is instructive to scale the emission frequency with the lattice parameter (a) and the speed of light in vacuum (c) according to $\tilde{\omega} = \omega a / 2\pi c$. At fixed emission frequency ω , we compare quantum

dots of the same size and hence with identical emission properties. Figure 4 displays the measured total decay rate as a function of scaled frequency for three fixed emission frequencies corresponding to large (ω_L), medium (ω_M) and small (ω_S) sized quantum dots. The experimental data are compared to the calculated density of states (DOS). The DOS is the unit-cell averaged LDOS weighted by all Bloch mode functions, and approximates the LDOS well in the weakly photonic limit $\tilde{\omega} \ll 1$, where spatial variations are less pronounced. We compare the data to the DOS because it can readily be calculated, whereas LDOS calculations are a tremendous computational task²⁰ and existing simulations are currently under debate²³. At low scaled frequencies ($\tilde{\omega} < 0.65$), the photonic crystal is effectively a homogeneous medium, and we measure a non-photonic reference decay rate γ_h . As a is increased, the decay rate is reduced for all quantum dot sizes, and a broad band of inhibited emission is observed in agreement with previous continuous-wave measurements²⁴. The inhibition is found in the pseudogap region where the DOS is strongly altered. The rapid drop in the measured decay rate is qualitatively described by the DOS. At even higher frequencies ($\tilde{\omega} > 0.85$), the decay rates are strongly enhanced compared to the reference decay rate over a wide frequency range. In this range, the calculated DOS also reveals enhancements, but this is the strongly photonic regime where the DOS is no longer an adequate description, and local variations described by the LDOS are crucial.

In most previous studies on emission in photonic crystals, dye molecules have been used as light sources. Recent calculations confirm that earlier claims of lifetime effects were not caused by the photonic structure²⁵. A fivefold inhibition of the radiative decay rate was concluded from a continuous-wave measurement of the integrated emission performed with a very low quantum efficiency light source²⁴. In contrast, a photonic effect on the emission lifetime is complementary and requires high quantum efficiency of the light source. Direct comparison between the present time-resolved and previous continuous-wave experiments is not warranted. In the present measurements, the complex decay curves imply that an averaged lifetime is derived. Interpretation of long-time components of the decay curves in terms of local variations in the lifetime is restricted by fluctuations in the non-radiative decay rate²². Furthermore, the non-radiative decay reduces the quantum efficiency. Therefore, a less pronounced modification of the total decay rate is observed in our time-resolved experiment, consistent with the continuous-wave experiment that provides an upper bound. Our lifetime measurements are distinctive because they provide actual control of the population of the excited states, which was not achieved previously.

Our findings of the modified lifetime in a photonic crystal can be compared to measurements of the Purcell effect in high-finesse cavities. Strong enhancement by up to a factor of five^{26,27} and vacuum-Rabi oscillations²⁸ have been observed on resonance in cavity quantum electrodynamic experiments. However, this enhancement is limited to a narrow bandwidth of $\Delta\omega/\omega \approx 10^{-4}$, determined by the linewidth of the cavity. Inhibition of spontaneous emission by a factor of ten in a bandwidth of 10^{-3} has been measured off-resonance in sidewall-coated microcavities²⁷. Our measurements of both inhibition and enhancement using photonic crystals extend over an unprecedentedly wide bandwidth of more than 10%, which is 100 times wider than in cavities. Furthermore, a serious restriction in a cavity is that control is limited to the small volume of a cavity (typically of the order of wavelength cubed), whereas in the current experiment the quantum dots were located throughout the large volume of the whole photonic crystal ($\sim 1\text{ mm}^3$). This clearly demonstrates the advantage of photonic crystals for broadband and large-volume applications with potential for efficient light sources and photon harvesting. We expect our results to open novel avenues for controlling quantum coherence in condensed matter environments¹⁸. □

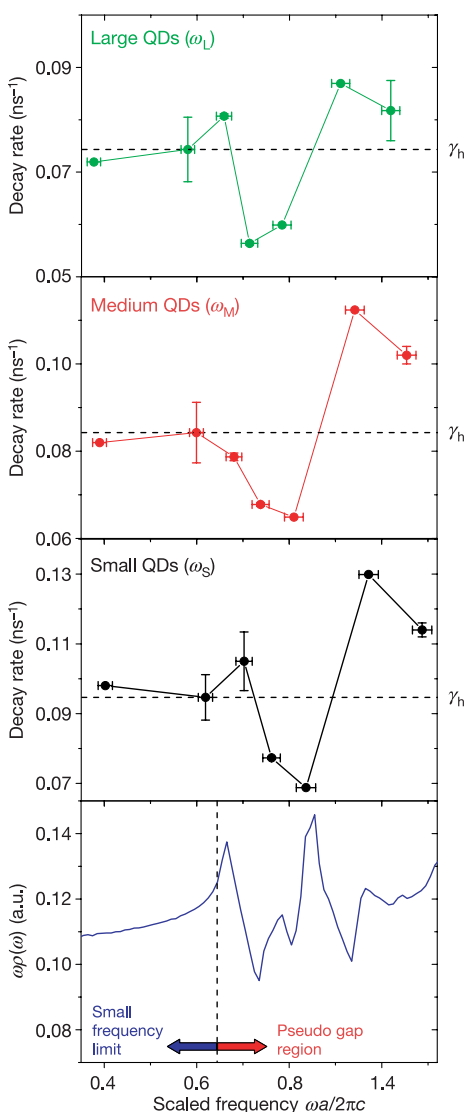


Figure 4 Measured decay rate at three fixed frequencies, ω_L , ω_M and ω_S , corresponding to large, medium and small sized quantum dots, when varying the lattice parameter a . The plotted curve (bottom) is the emission frequency ω multiplied by the calculated DOS $\rho(\omega)$ for a titania inverse opal, which is proportional to the decay rate according to Fermi's golden rule. Also indicated (dashed horizontal lines) are the decay rates γ_h in the limit of a homogeneous medium. The error bars indicate the standard error of the measurement. The vertical error bars are obtained by averaging over different samples with the same lattice parameters, and by probing different parts of the same sample, while the horizontal error bars are from measurements of the lattice parameter with a scanning electron microscope.

Methods

Sample preparation

The experiment was performed with titania inverse opals that consist of a face-centred cubic (f.c.c.) structure of air spheres in a titania backbone material. The inverse opals were fabricated by inverting template assisted self-assembled structures¹⁶. Typical dimensions of the samples were $2 \times 2 \times 0.3 \text{ mm}^3$. The high quality of the crystals is quantified by the scattering mean free path of light propagation, which was 15–20 μm . In total, ten different samples were studied with lattice parameters ranging from $a = 240 \text{ nm}$ to 650 nm . The photonic crystals were infiltrated for 24 h with a mixture of 50% chloroform and 50% butanol containing $10^{-7} \text{ mol litre}^{-1}$ ZnSe-coated CdSe colloidal quantum dots²⁹. Afterwards the samples were rinsed for one minute in chloroform and dried. The quantum dots had a size dispersity of 5% around an average diameter of 4.5 nm. The infiltration led to a distribution of quantum dots on the surfaces of the air spheres inside the inverse opal with a concentration of less than 10 quantum dots per air sphere. This is sufficiently low to avoid energy transfer between quantum dots. The ensemble of quantum dots in the solution had a lifetime of around 20 ns and a quantum efficiency of $50\% \pm 5\%$; the latter was measured by comparing absorption and emission to a R6G dye standard. To minimize oxidation and contamination, all sample preparation and handling was carried out in a nitrogen-purged glove box, and the optical measurements were performed in a sealed chamber under 1.7 mbar nitrogen.

Experimental set-up

Optical excitation of the quantum dots was performed with either a continuous-wave argon laser (488 nm) or a pulsed diode laser (100 ps, 440 nm). The former was used for the spectral measurements and the latter for time-resolved experiments. The excitation beam was coupled through a single-mode fibre and illuminated the photonic crystal. The fibre and the sample were mounted on the same rotational stage, which allowed changing of the detection angle while keeping the excitation volume constant. The spontaneously emitted light was detected from the surface facing away from the excitation beam at an angle α relative to the (111) planes of the f.c.c. lattice. The light was collected in a cone of $\sim 8^\circ$ around the detection angle, dispersed with a spectrometer (2 nm resolution), and detected with a fast photomultiplier. Time-correlated single-photon counting was implemented to measure the arrival time of the emitted photons. Careful measurements on reference samples ruled out the presence of energy transfer³⁰, reabsorption, and other non-photonic effects on the lifetime.

Data analysis

The calculated curves in Fig. 1c were obtained with only experimental data from separate reflectivity experiments as input to the theory, combined with band structure calculations²¹. Furthermore, the lattice parameters of the photonic crystals were taken from scanning electron microscope measurements.

The decay curves in Fig. 2 were obtained by binning the arrival times of the single photons detected with the photomultiplier. Contributions from dark counts in the photomultiplier were measured and subtracted from the data. The first 5 ns were discarded from each decay curve, as defect-related luminescence of titania was contributing in this range. The signal beyond 5 ns was unaffected by titania luminescence, and the decay until $\sim 20 \text{ ns}$ was modelled with a single exponential decay from which the lifetime was extracted. Whereas strong lifetime effects were observed on the efficient quantum dots, we observe no systematic change of the lifetime from the low quantum efficiency titania luminescence.

The calculated lifetime of quantum dot emission in a homogeneous medium in Fig. 3 is given by $\tau_h = C/\omega^3$. Here $C = 3\hbar\epsilon_0\pi^2c/n\bar{d}^2$ contains Planck's constant $\hbar$, the vacuum permittivity ϵ_0 , the speed of light in vacuum c , the refractive index n , and the mean transition-dipole moment $\bar{d}$ averaged over the orientation and size distribution of the quantum dots. We assume an average refractive index of $n = 1.17$ and arrive at a mean dipole moment of 3.0 atomic units.

Received 25 February; accepted 16 June 2004; doi:10.1038/nature02772.

- Yablonovitch, E. Inhibited spontaneous emission in solid-state physics and electronics. *Phys. Rev. Lett.* **58**, 2059–2062 (1987).
- Painter, O. *et al.* Two-dimensional photonic band-gap defect mode laser. *Science* **284**, 1819–1821 (1999).
- Smith, C. J. M. *et al.* Near-infrared microcavities confined by two-dimensional photonic bandgap crystals. *Electron. Lett.* **35**, 228–230 (1999).
- Colombelli, R. *et al.* Quantum cascade surface-emitting photonic crystal laser. *Science* **302**, 1374–1377 (2003).
- Noda, S. *et al.* Polarization mode control of two-dimensional photonic crystal laser by unit cell structure design. *Science* **293**, 1123–1125 (2001).
- Kim, J., Benson, O., Kan, H. & Yamamoto, Y. A single-photon turnstile device. *Nature* **397**, 500–503 (1999).
- Michler, P. *et al.* A quantum dot single-photon turnstile device. *Science* **290**, 2282–2285 (2000).
- Kuhn, A., Hennrich, M. & Rempe, G. Deterministic single-photon source for distributed quantum networking. *Phys. Rev. Lett.* **89**, 067901 (2002).
- Grätzel, M. Photoelectrochemical cells. *Nature* **414**, 338–344 (2001).
- Scully, M. O. & Zubairy, M. S. *Quantum Optics* (Cambridge Univ. Press, Cambridge, 1997).
- Loudon, R. *The Quantum Theory of Light* (Oxford Univ. Press, New York, 2000).
- Bykov, V. P. Spontaneous emission from a medium with a band spectrum. *Sov. J. Quant. Electron.* **4**, 861–871 (1975).
- Lambropoulos, P., Nikolopoulos, G. M., Nielsen, T. R. & Bay, S. Fundamental quantum optics in structured reservoirs. *Rep. Prog. Phys.* **63**, 455–503 (2000).
- John, S. & Quang, T. Spontaneous emission near the edge of a photonic band gap. *Phys. Rev. A* **50**, 1764–1769 (1994).
- John, S. Strong localization of photons in certain disordered dielectric superlattices. *Phys. Rev. Lett.* **58**, 2486–2489 (1987).

- Wijnhoven, J. E. G. J. & Vos, W. L. Preparation of photonic crystals made of air spheres in titania. *Science* **281**, 802–804 (1998).
- Vahala, K. J. Optical microcavities. *Nature* **424**, 839–846 (2003).
- Mabuchi, H. & Doherty, A. C. Cavity quantum electrodynamics: coherence in context. *Science* **298**, 1372–1377 (2002).
- Sprink, R., van Tiggelen, B. A. & Lagendijk, A. Optical emission in periodic dielectrics. *Europhys. Lett.* **35**, 265–270 (1996).
- Busch, K. & John, S. Photonic band gap formation in certain self-organizing systems. *Phys. Rev. E* **58**, 3896–3908 (1998).
- Koenderink, A. F. & Vos, W. L. Light exiting from real photonic band gap crystals is diffuse and strongly directional. *Phys. Rev. Lett.* **91**, 213902 (2003).
- Fisher, B. R., Eisler, H.-J., Stott, N. E. & Bawendi, M. G. Emission intensity dependence and single-exponential behavior in single colloidal quantum dot fluorescence lifetimes. *J. Phys. Chem. B* **108**, 143–148 (2004).
- Wang, R., Wang, X.-H., Gu, B.-Y. & Yang, G.-Z. Local density of states in three-dimensional photonic crystals: Calculation and enhancement effects. *Phys. Rev. B* **67**, 155114 (2003).
- Koenderink, A. F., Bechger, L., Schriemer, H. P., Lagendijk, A. & Vos, W. L. Broadband fivefold reduction of vacuum fluctuations probed by dye in photonic crystals. *Phys. Rev. Lett.* **88**, 143903 (2002).
- Li, Z. Y. & Zhang, Z. Q. Weak photonic band gap effect on the fluorescence lifetime in three-dimensional colloidal photonic crystals. *Phys. Rev. B* **63**, 125106 (2001).
- Gérard, J. M. *et al.* Enhanced spontaneous emission by quantum boxes in a monolithic optical microcavity. *Phys. Rev. Lett.* **81**, 1110–1113 (1998).
- Bayer, M. *et al.* Inhibition and enhancement of the spontaneous emission of quantum dots in structured microresonators. *Phys. Rev. Lett.* **86**, 3168–3171 (2001).
- Hood, C. J., Chapman, M. S., Lynn, T. W. & Kimble, H. J. Real-time cavity QED with single atoms. *Phys. Rev. Lett.* **80**, 4157–4160 (1998).
- Dabbousi, B. O. *et al.* (CdSe)/ZnS core-shell quantum dots: synthesis and characterization of highly luminescent nanocrystallites. *J. Phys. Chem. B* **101**, 9463–9475 (1997).
- Crooker, S. A., Hollingsworth, J. A., Tretyak, S. & Klimov, V. I. Spectrally resolved dynamics of energy transfer in quantum-dot assemblies: towards engineered energy flows in artificial materials. *Phys. Rev. Lett.* **89**, 186802 (2002).

Acknowledgements We thank L. Woldering for sample preparation, F. Koenderink for DOS calculations, A. Mosk and J. Kelly for discussions, and A. Lagendijk for support. This work is part of the research programme of the Stichting voor Fundamenteel Onderzoek der Materie (FOM), which is financially supported by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.L. (p.lodahl@utwente.nl).

Atomic-scale imaging of nanoengineered oxygen vacancy profiles in SrTiO₃

David A. Muller^{1,2}, Naoyuki Nakagawa^{1,3}, Akira Ohtomo^{1,4}, John L. Grazul^{1,2} & Harold Y. Hwang^{1,3,5}

¹Bell Laboratories, Lucent Technologies, Murray Hill, New Jersey 07974, USA

²School of Applied and Engineering Physics, Cornell University, Ithaca, New York 14853, USA

³Department of Advanced Materials Science, University of Tokyo, Kashiwa, Chiba 277-8561, Japan

⁴Institute of Materials Research, Tohoku University, Sendai, 980-8577, Japan

⁵Japan Science and Technology Agency, Kawaguchi, 332-0012, Japan

At the heart of modern oxide chemistry lies the recognition that beneficial (as well as deleterious) materials properties can be obtained by deliberate deviations of oxygen atom occupancy from the ideal stoichiometry^{1,2}. Conversely, the capability to control and confine oxygen vacancies will be important to realize the full potential of perovskite ferroelectric materials, varistors and field-effect devices^{3,4}. In transition metal oxides, oxygen vacancies are generally electron donors, and in strontium titanate (SrTiO₃) thin films, oxygen vacancies (unlike impurity dopants) are particularly important because they tend to retain high carrier mobilities, even at high carrier densities⁵. Here we report the successful fabrication, using a pulsed laser deposition technique, of SrTiO₃ superlattice films with oxygen doping profiles

that exhibit subnanometre abruptness. We profile the vacancy concentrations on an atomic scale using annular-dark-field electron microscopy and core-level spectroscopy, and demonstrate absolute detection sensitivities of one to four oxygen vacancies. Our findings open a pathway to the microscopic study of individual vacancies and their clustering, not only in oxides, but in crystalline materials more generally.

The introduction of mobile charge carriers through the controlled placement of ionizable impurities is central to the production of semiconductor devices⁶. Recently, SrTiO₃ devices have been produced using oxygen vacancies as dopants, as their carrier mobility can reach $10^4 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ at low temperatures^{5,7}. Although ultimately a further reduction in device size of another two orders of magnitude is desirable, we are faced with two questions. First, how small can a non-stoichiometric layer be made and remain stable against diffusion; and second, how to measure what are essentially macroscopically low vacancy concentrations on an atomic length scale buried inside a material. Solutions to the latter question have implications well beyond oxide electronics. As an example, charge disproportionation in strongly correlated oxides is heavily influenced by the length scale over which vacancies cluster, if at all⁸. If vacancy clustering correlations are shorter than the interaction lengths for the electron gas, the crystal might be approximated as a homogenous medium (as is largely the case for semiconductors). If the opposite occurs, the motion of droplets of a phase-separated electron liquid may be modulated or pinned.

In order to measure the oxygen diffusion vacancy distribution that occurred during growth (in addition to the efficiency of the self-doping of vacancies), we grew SrTiO_{3- δ} films with precisely controlled thickness and known electrical properties. Typical useful vacancy ranges are 1–4%, which are too low to be detected by the recently proposed ‘direct’ imaging in a transmission electron microscope that has been corrected for spherical aberration⁹. Instead, we have investigated ‘indirect’ methods with an order of magnitude increase in sensitivity, and with unit-cell resolution, that

are more quantitative and sensitive to the electronic properties of SrTiO₃. Here we present not only a method to image as few as 1–4 oxygen vacancies and hence the most intimate details of vacancy clustering, but also demonstrate that artificial, vacancy-engineered structures can be produced with nanometre-scale abruptness. The limiting factor to this abruptness is the clustering of vacancies over a few unit cells length for our growth conditions.

SrTiO_{3- δ} homoepitaxial thin films were grown using pulsed laser deposition from a single-crystal SrTiO₃ target. This deposition technique is well suited for these studies, because it is one of the most kinetic methods for deposition. This allows significantly enhanced diffusion of the deposited adatoms while maintaining a relatively low substrate temperature, minimizing thermal diffusion. Films were grown on TiO₂-terminated (001) SrTiO₃ substrates,

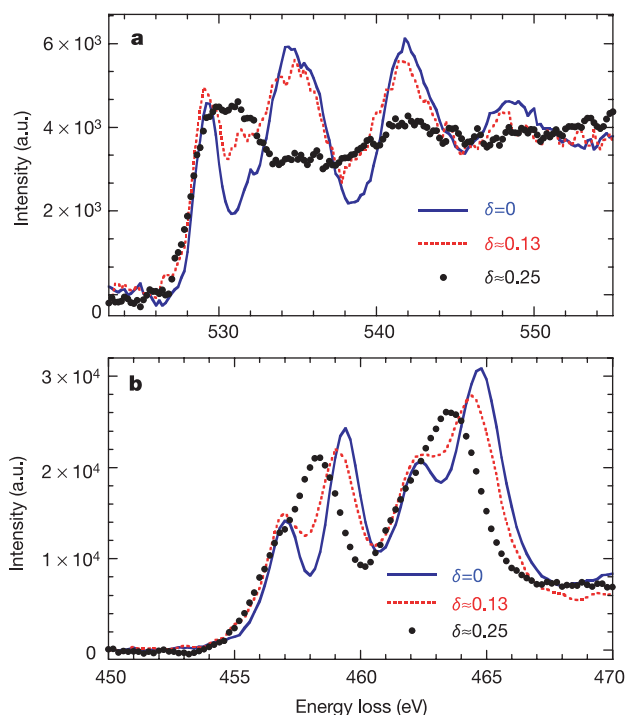


Figure 1 Electron energy-loss spectra (EELS) for oxygen-deficient SrTiO_{3- δ} , for $\delta \approx 0$, 0.13 and 0.25. **a**, The O-K edge fine structure is sensitive to O–O ordering, and damps out as the vacancy concentration increases. **b**, The Ti-L edge shifts from a 4⁺ to 3⁺ valence with increasing [V_O], as the oxygen vacancies are electron donors.

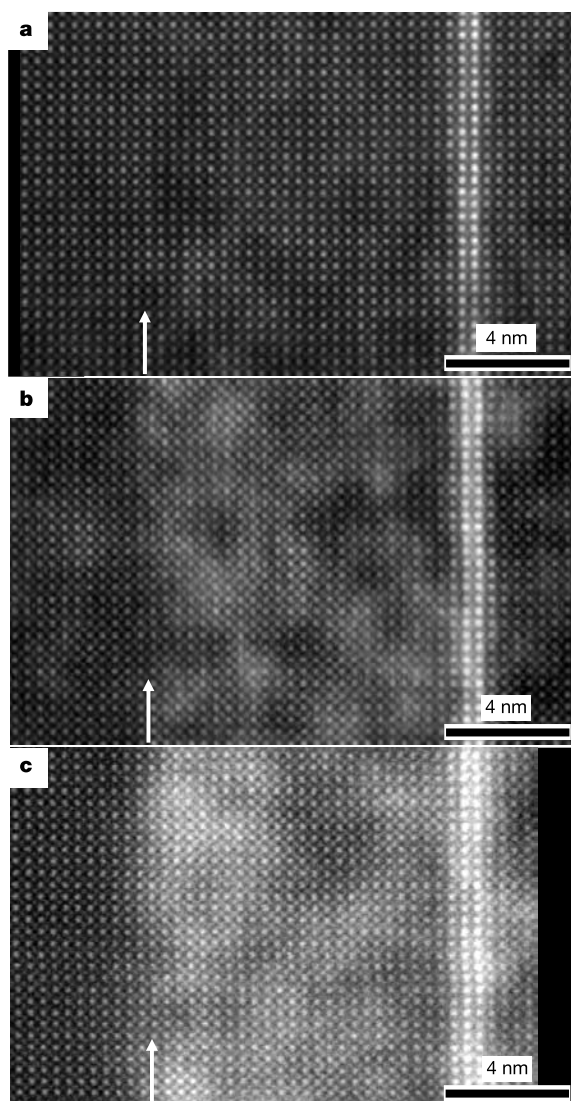


Figure 2 25 unit cells of oxygen-deficient SrTiO_{3- δ} ($\delta \approx 0.13$) are grown on bulk SrTiO₃ (left of white arrow) and capped with a two-layer-thick LaTiO₃ marker layer (right) and more SrTiO_{3- δ} . **a**, High-angle annular dark field (HAADF) imaging is sensitive mostly to atomic number and shows only the La marker layer. The absence of an intensity change across the SrTiO₃/SrTiO_{3- δ} interface confirms no preferential thinning of the SrTiO_{3- δ} layer. **b**, Low-angle annular dark field (LAADF) imaging also records dechannelling from the strain field surrounding the O vacancies, so the oxygen-deficient layer is now also visible—this sample is 10 nm thick. **c**, LAADF image in a ~30-nm-thick sample. Data from a JEOL 2010F STEM with a 1.05 mm spherical aberration coefficient, at 197 kV with a 10 mrad convergence semiangle, and a 50 mrad ADF inner angle for HAADF and a 25 mrad inner angle for LAADF.

characterized by atomically flat terraces separated by unit cell steps, using a KrF excimer laser at a repetition rate of 4 Hz and a laser fluence at the target surface of $\sim 3 \text{ J cm}^{-2}$. The growth temperature was 750–800 °C, with an oxygen partial pressure p_{O_2} ranging from 5×10^{-5} to 5×10^{-7} torr. Unit cell reflection high-energy electron diffraction (RHEED) intensity oscillations were observed throughout the growth (15–25 pulses per unit cell), and were used to calibrate the number of layers grown. After growth, the samples were cooled at $50^\circ \text{C min}^{-1}$ in the same p_{O_2} as the last stage of growth. For detailed characterization we used a scanning transmission electron microscope (STEM), which can detect individual impurity atoms with atomic resolution¹⁰. Here a 200-kV electron beam is focused down to a spot that is less than 0.2 nm in diameter^{11,12}, and is scanned across a cross-sectioned sample that has been thinned down to electron transparency (by tripod polishing, followed by low-energy, low-angle ion milling on a liquid-nitrogen cooled stage to minimize damage and vacancy redistribution)¹¹. The transmitted electrons can either be analysed by an on-axis electron energy-loss spectrometer, allowing atomic-scale spectroscopy^{13–15}, or collected by an annular dark field detector that surrounds the spectrometer. This is an efficient division of signals, as the inelastic scattering used for spectroscopy is strongly peaked at small scattering angles ($< 20 \text{ mrad}$) while the elastic/phonon scattering that forms the annular dark field (ADF) images dominates at higher angles¹⁶.

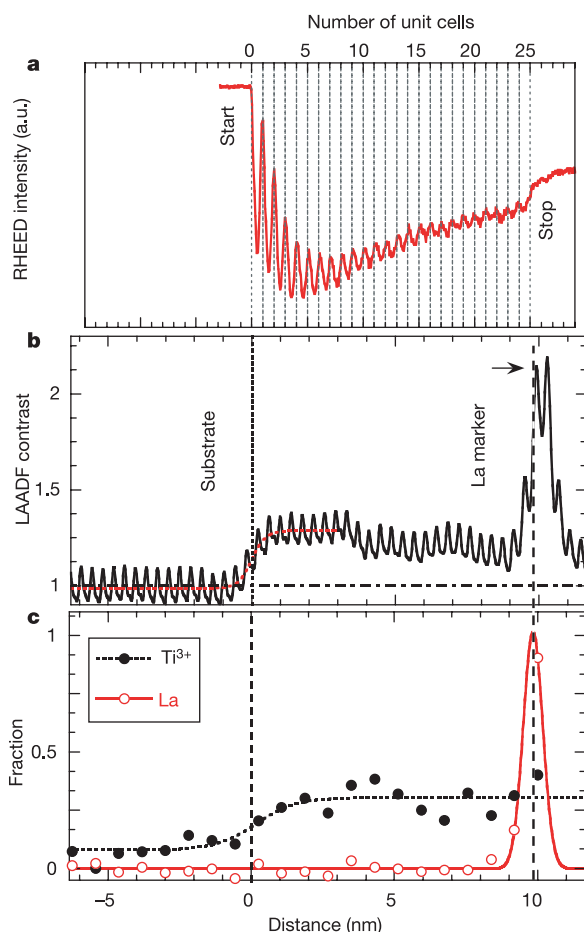


Figure 3 Quantitative line profiles through the oxygen-deficient layer. **a**, Reflection high-energy electron diffraction (RHEED) pattern during growth shows that 25 unit cells are grown. **b**, LAADF line profile—the original interface at 0 nm is determined by counting 25 unit cells back from the La marker. **c**, EELS profile showing the Ti^{3+} fraction (from the Ti-L edge) and the simultaneously recorded La-M edge. Each EELS spectrum was recorded with a 0.16 nm spot size, 15 pA beam current and 10 s dwell time, and the Ti^{3+} fraction determined from a least-squares fit to reference spectra as described in ref. 23.

Electron energy-loss spectroscopy (EELS) provides information on both the O and Ti states from the O-K ($\text{O } 1s \rightarrow 2p$) and Ti-L ($\text{Ti } 2p \rightarrow 3d$) core edges. The Ti/O ratio could be determined directly from the ratios of integrated O-K and Ti-L edge intensities, but this is not practical for such low oxygen vacancy concentrations [V_{O}] as the errors in background subtraction are usually a few per cent. Instead we rely on the shape of the EELS edge, which reflects the underlying electronic structure—see appendix A of ref. 17. Figure 1a shows that the O-K edge fine structure (which is proportional to the unoccupied O p -density of states in the presence of a core hole) ‘washes out’ with increasing [V_{O}] for a sequence of films, and even at $\delta \approx 0.13$ (4% V_{O}), changes are visible. This trend is consistent with cluster simulations of grain boundaries in SrTiO_3 with reduced oxygen coordination shells¹⁸. The nominal oxygen stoichiometry was estimated using the Hall effect to extract the free carrier density, and assuming two free carriers for each oxygen vacancy, as has been previously reported¹⁹ (although at high vacancy concentrations the number of free electrons per vacancy may be reduced²⁰). Even more useful is the Ti-L edge (Fig. 1b), which shows distinctly different spectra for the Ti^{3+} and Ti^{4+} formal valences^{21,22}. Each oxygen vacancy nominally ‘transfers’ two electrons to the Ti d band. As there are three times as many O sites as Ti, this is an amplification factor of six in fractional detection sensitivity. The formal Ti valence can be extracted by a multiple least-squares fit to Ti^{3+} and Ti^{4+} reference spectra²³ with a detection limit of $\sim \delta = 0.05$ (that is, about 1% V_{O}). The EELS spectra provide chemical and electronic information on buried atoms that are in a bulk-like environment¹¹.

In order to study the interface between the stoichiometric substrate and an oxygen-deficient film, 25 unit cells of $\text{SrTiO}_{3-\delta}$ were grown (800 °C, $p_{\text{O}_2} = 10^{-6}$ torr—conditions that we found did not alter the oxygen stoichiometry of the substrate surface), and then capped with 2 unit cells of LaTiO_3 to serve as a marker. This was followed by further $\text{SrTiO}_{3-\delta}$ growth. The high-angle annular dark field (HAADF) signal (50–250 mrad) scales approximately as $Z^{1.7}$, where Z is the atomic number (refs 24, 25). Thus the HAADF image (Fig. 2a) is dominated by the (high- Z) cation sites, and neither the oxygen atoms nor vacancies are visible (the interface is marked by an arrow). In addition to atomic number, the low-angle annular dark field (LAADF) signal at 25–50 mrad is very sensitive to

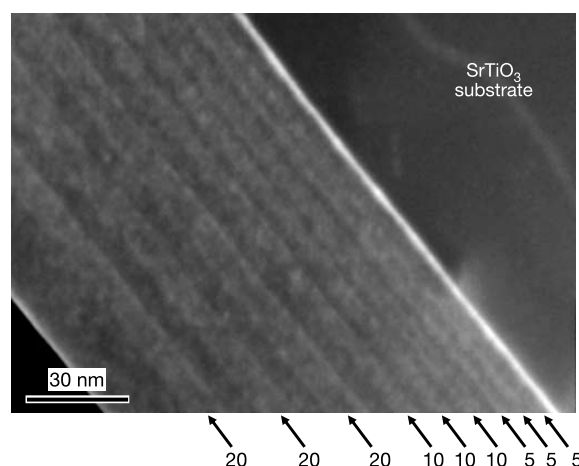


Figure 4 A LAADF image of an oxygen-modulated superlattice grown on SrTiO_3 . The growth sequence is $3 \times (5 \text{ unit cells of } \text{SrTiO}_{3-\delta}, 5 \text{ of } \text{SrTiO}_3), 3 \times (10 \text{ unit cells of } \text{SrTiO}_{3-\delta}, 10 \text{ of } \text{SrTiO}_3), 3 \times (20 \text{ unit cells of } \text{SrTiO}_{3-\delta}, 20 \text{ of } \text{SrTiO}_3), 50 \text{ of } \text{SrTiO}_3$. The sample was grown at 760 °C, with p_{O_2} alternating between 5×10^{-7} and 5×10^{-5} torr, corresponding to $\text{SrTiO}_{3-\delta}$ and SrTiO_3 growth, respectively, with approximately 2 min intervals between growth at the different p_{O_2} levels. The initial starting surface was reduced by the low starting pressure under pre-annealing conditions at higher temperature, appearing as a strong bright line at the interface. By 10^{-6} torr, this line is eliminated, preserving a stoichiometric surface (such as the sample shown in Fig. 2).

any effect that leads to a dechannelling of the incident electron beam—such as phonons or strain fields from point defects^{26,27}. Figure 2b shows that the LAADF signal can detect the strain fields from oxygen vacancies through their distortion of the adjacent cation sites and subsequent dechannelling of the electron beam from the cation columns. The strain field is expected to decay below the dechannelling detection limit by the next-nearest neighbour sites (and be most strongly peaked on the cation nearest neighbours)²⁶, so at least a unit-cell resolution should be expected for a LAADF image of the vacancy distribution. Figure 2c shows a thicker region (~30 nm) that averages over more vacancies in the beam direction.

How far has the oxygen diffused during the growth of the test structure? Fig. 3a shows the RHEED pattern used to grow the 25 layers of oxygen-deficient SrTiO_{3-δ}. By counting back from the La marker layer in the LAADF STEM image (Fig. 2b), the original substrate interface can be located, and we see that vacancies have not diffused down into the substrate. This is most readily apparent in Fig. 3b, which shows the average line profile from Fig. 2b. From a gaussian fit to the interface profile, the average diffusional/cluster broadening has at most a root-mean-square width of 0.38 ± 0.1 nm (about one unit cell). This is also an upper limit on the spatial resolution of the LAADF method of imaging vacancies. Figure 3b shows that the electrical width from the Ti EELS signal is ~1 nm, consistent with the measured screening length in delta-doped SrTiO₃ (ref. 23). The average percentage of Ti³⁺ in the oxygen-deficient region was $25 \pm 5\%$, which would correspond to an oxygen deficiency of $\delta = 0.13 \pm 0.03$, within experimental error of the nominal $\delta \approx 0.13 \pm 0.06$ (as deduced from Hall effect measurements from samples without the marker layer grown under the same nominal conditions). Images of the $\delta \approx 0.25$ sample (deduced by Hall measurements on this sample) are given in Supplementary Fig. 1. Residual surface states are probably responsible for the small (~5%) Ti³⁺ offset in the substrate. This could add a systematic error of $\delta \approx 0.025$ to the EELS quantification in thin samples.

By matching the contrast changes in the simultaneously recorded bright-field image with multislice simulations, we estimate the sample in Fig. 2b to be 10 ± 2 nm thick. As the average oxygen vacancy concentration is 4% ($\delta \approx 0.13$), this implies that the average number of oxygen vacancies per column is 3 ± 1 . The signal to noise ratio (SNR) for the EELS line scan of Fig. 3c gives an estimate of the absolute detection sensitivity for quantification of the Ti L edge that is 2–4 oxygen vacancies and 4–8 electrons on the Ti sites. The EELS signal can be calibrated from reference spectra, but the LAADF signal should be thickness dependent^{26,27} and requires an independent means of calibration (such as the EELS signal). The averaged LAADF profile in Fig. 3b has a noise floor of 0.3% oxygen vacancies (or 0.2 vacancies per column for this sample thickness). In the actual image, however, where less averaging is possible, the noise fluctuations from the surface layers lead to a larger uncertainty of ± 0.7 vacancy per projected column of unit-cell width. This is obviously not sufficient to detect a single isolated vacancy with a high degree of confidence, but sufficient to discriminate a cluster as small as 2 vacancies from the noise levels. This limitation is probably from the surface preparation of the sample for TEM work, and until a method can be found to prepare perfect surfaces, it will remain the limiting factor for any TEM method, irrespective of the microscope's performance. LAADF (which should work for any crystalline material) has a unique advantage in this respect: the channelling contrast that forms the LAADF signal takes a few unit cells to build up to peak intensity²⁸ so the contribution from entrance-surface defects is greatly reduced compared to bulk defects. Exit-surface contrast can also be minimized by choosing thicker samples where dechannelling has occurred.

With the image intensity fluctuations quantified, we find that the intensity fluctuations are larger than would be expected from

Poisson statistics, suggesting that some vacancy clustering has occurred. Autocorrelation functions of the oxygen-deficient region have a half-width of 1.8 ± 0.1 nm (a cluster diameter of 4.6 ± 0.3 unit cells), while the autocorrelation functions measured on the stoichiometric substrate where the contrast fluctuations are much less intense (such as the left half of Fig. 2b) have a half-width of 0.3 ± 0.1 nm (0.8 ± 0.3 unit cells), which is essentially uncorrelated. These residual fluctuations are presumably the result of surface damage during sample preparation. Figure 2c shows a thicker cross-section (~30 nm) that averages over more vacancies in the beam direction, removing the high spatial frequencies from the correlation function. The region in Fig. 2b should be too thin for this averaging to be significant.

Although vacancy clustering has not been previously reported in SrTiO₃ thin films, their presence is reasonable in view of the fact that there appear to be thermodynamic limits to the production of a uniform distribution of vacancies. Recent work²⁹ suggests that “it may be impossible to homogeneously reduce single crystals of SrTiO₃ and describe their behavior in terms of point defect chemistry”. Instead, the oxygen-deficient region may shear and collapse, with non-perovskite Ti-rich phases appearing on the reduced surfaces³⁰. It is important to consider these issues when considering the ultimate limits to controlling the stoichiometry profile in a thin film structure. In order to experimentally test these limits, we grew an oxygen-vacancy-modulated superlattice structure, where layers of stoichiometric and oxygen-deficient SrTiO₃ of differing thicknesses are grown in repeated patterns (Fig. 4). Not only is it possible to grow low [V_O]-doped material on high [V_O]-doped material to produce alternating layers of low and high conductivity, but it is also possible to grow relatively thick structures where the dopant profiles can be engineered on nanometre length scales. Although Fig. 2 shows that vacancy diffusion into the stoichiometric substrate during growth is limited to one unit cell or less, the five-unit-cell vacancy cluster correlation length seems to reduce the abruptness of the subsequent interfaces. Supplementary Fig. 2 shows that periodically layered structures must be at least three unit cells in order to survive.

The present work shows that it is possible to grow, image and quantify vacancy distributions at a microscopic level, and so opens the way to systematic studies of vacancy clustering and deactivation in oxides. Modelling of the O-K near-edge structure may allow physical structures of these defect–defect interactions at high concentrations to be investigated. □

Received 7 March; accepted 14 June 2004; doi:10.1038/nature02756.

1. Majewski, P. BiSrCaCuO high-Tc superconductors. *Adv. Mater.* **6**, 460–469 (1994).
2. Shimakawa, Y., Kubo, Y., Manako, T. & Igarashi, H. Variation in T_c and carrier concentration in Ti-based superconductors. *Phys. Rev. B* **40**, 11400–11402 (1989).
3. Greuter, F. & Blatter, G. Electrical properties of grain boundaries in polycrystalline compound semiconductors. *Semicond. Sci. Technol.* **5**, 111–137 (1990).
4. Ahn, C. H., Triscone, J.-M. & Mannhart, J. Electric field effect in correlated oxide systems. *Nature* **424**, 1015–1018 (2004).
5. Tufte, O. N. & Chapman, P. W. Electron mobility in semiconducting strontium titanate. *Phys. Rev.* **155**, 796–802 (1967).
6. Sze, S. M. *Semiconductor Devices: Physics and Technology* (John Wiley, New York, 1986).
7. Pellegrino, L., Palleschi, L., Marré, D., Bellingeri, E. & Siri, A. S. Fabrication of submicron-scale SrTiO_{3-δ} devices by an atomic force microscope. *Appl. Phys. Lett.* **81**, 3849–3851 (2002).
8. Burg, J., Mayr, M., Martin-Mayor, V., Moreo, A. & Dagotto, E. Colossal effects in transition metal oxides caused by intrinsic inhomogeneities. *Phys. Rev. Lett.* **87**, 277202 (2001).
9. Jia, C. L., Lentzen, M. & Urban, K. Atomic-resolution imaging of oxygen in perovskite ceramics. *Science* **299**, 870–873 (2003).
10. Voyles, P. M., Muller, D. A., Grazul, J. L., Citrin, P. H. & Gossmann, H.-J. L. Atomic-scale imaging of individual dopant atoms and clusters in highly n-type bulk Si. *Nature* **416**, 826–829 (2002).
11. Kaiser, U., Muller, D. A., Grazul, J., Chuvulin, A. & Kawasaki, M. Direct observation of defect-mediated cluster nucleation. *Nature Mater.* **1**, 102–105 (2002).
12. Muller, D. A. & Grazul, J. Optimizing the environment for sub-0.2 nm scanning transmission electron microscopy. *J. Elec. Microsc.* **50**, 219–226 (2001).
13. Batson, P. E. Simultaneous STEM imaging and electron energy-loss spectroscopy with atomic column sensitivity. *Nature* **366**, 727–728 (1993).
14. Browning, N. D., Chisholm, M. M. & Pennycook, S. J. Atomic-resolution chemical analysis using a scanning transmission electron microscope. *Nature* **366**, 143–146 (1993).
15. Muller, D. A. *et al.* The electronic structure at the atomic scale of ultra-thin gate oxides. *Nature* **399**, 758–761 (1999).

16. Egerton, R. F. *Electron Energy Loss Spectroscopy in the Electron Microscope* (Plenum, New York, 1996).
17. Zhang, Z., Sigle, W. & Rühle, M. Atomic and electronic characterization of the $a[100]$ dislocation core in SrTiO_3 . *Phys. Rev. B* **66**, 094108 (2002).
18. Browning, N. D., Moltaji, H. O. & Buban, J. P. Investigation of three-dimensional grain-boundary structures in oxides through multiple-scattering analysis of spatially resolved electron-energy-loss spectra. *Phys. Rev. B* **58**, 8289–8300 (1998).
19. Yamada, H. & Miller, G. R. Point defects in reduced strontium titanate. *J. Solid State Chem.* **6**, 169–177 (1973).
20. Gong, W. *et al.* Oxygen-deficient SrTiO_{3-x} , $x = 0.28, 0.17, 0.08$. Crystal growth, crystal structures, magnetic, and transport properties. *J. Solid State Chem.* **90**, 320–330 (1991).
21. Abbate, M. *et al.* Soft-x-ray-absorption studies of the location of extra charges induced by substitution in controlled-valence materials. *Phys. Rev. B* **44**, 5419–5422 (1991).
22. Ohtomo, A., Muller, D. A., Grazul, J. L. & Hwang, H. Y. Epitaxial growth and electronic structure of LaTiO_x films. *Appl. Phys. Lett.* **80**, 3922–3925 (2002).
23. Ohtomo, A., Muller, D. A., Grazul, J. L. & Hwang, H. Y. Artificial charge-modulation in atomic-scale perovskite titanate superlattices. *Nature* **419**, 378–380 (2002).
24. Howie, A. Image contrast and localized signal selection techniques. *J. Microsc.* **17**, 11–23 (1979).
25. Kirkland, E. J., Loane, R. F. & Silcox, J. Simulation of annular dark field STEM images using a modified multislice method. *Ultramicroscopy* **23**, 77–96 (1987).
26. Hillyard, S. E. & Silcox, J. Detector geometry, thermal diffuse scattering and strain effects in ADF STEM imaging. *Ultramicroscopy* **58**, 6–17 (1995).
27. Perovic, D. D., Rossow, C. J. & Howie, A. Imaging elastic strains in high-angle annular dark-field scanning-transmission electron microscopy. *Ultramicroscopy* **52**, 353–359 (1993).
28. Voyles, P. M., Muller, D. A. & Kirkland, E. J. Depth-dependent imaging of individual dopant atoms in silicon. *Microsc. Microanal.* **10**, 291–300 (2004).
29. Sztot, K., Speier, W., Carius, R., Zastrow, U. & Beyer, W. Localized metallic conductivity and self-healing during thermal reduction of SrTiO_3 . *Phys. Rev. Lett.* **88**, 075508 (2002).
30. Sztot, K. & Speier, W. Surfaces of reduced and oxidized SrTiO_3 from atomic force microscopy. *Phys. Rev. B* **60**, 5909–5926 (1999).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We acknowledge partial support from NEDO's International Joint Research Program. N.N. acknowledges partial support from QPEC, Graduate School of Engineering, University of Tokyo. D.A.M. and J.L.G. received partial support from the Cornell Center for Materials Research, a NSF materials research science and engineering centre.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.A.M. (davidm@ccmr.cornell.edu).

Vigorous exchange between the Indian and Atlantic oceans at the end of the past five glacial periods

Frank J. C. Peeters^{1,2}, Ruth Acheson³, Geert-Jan A. Brummer², Wilhelmus P. M. de Ruijter⁴, Ralph R. Schneider⁵, Gerald M. Ganssen¹, Els Ufkes¹ & Dick Kroon¹

¹Department of Paleocology and Paleoclimatology, Faculty of Earth and Life Sciences, Vrije Universiteit, de Boelelaan 1085, 1081 HV, Amsterdam, The Netherlands

²Department of Marine Chemistry and Geology (MCG), Royal Netherlands Institute for Sea Research (NIOZ), PO Box 59, 1790 AB Den Burg, Texel, The Netherlands

³School of GeoSciences, The University of Edinburgh, Grant Institute, The King's Buildings, Edinburgh EH9 3JW, UK

⁴Faculty of Physics and Astronomy, Institute for Marine and Atmospheric Research Utrecht, Utrecht University, Princetonplein 5, 3584 CC Utrecht, The Netherlands

⁵Département de Géologie et Océanographie, UMR 5805 EPOC, CNRS/Université de Bordeaux I, 33405 Talence, France

The magnitude of heat and salt transfer between the Indian and Atlantic oceans through 'Agulhas leakage' is considered important for balancing the global thermohaline circulation^{1–3}. Increases or reductions of this leakage lead to strengthening or weakening of the Atlantic meridional overturning and associated variation of North Atlantic Deep Water formation^{4–6}. Here we

show that modern Agulhas waters, which migrate into the south Atlantic Ocean in the form of an Agulhas ring, contain a characteristic assemblage of planktic foraminifera. We use this assemblage as a modern analogue to investigate the Agulhas leakage history over the past 550,000 years from a sediment record in the Cape basin. Our reconstruction indicates that Indian–Atlantic water exchange was highly variable: enhanced during present and past interglacials and largely reduced during glacial intervals. Coherent variability of Agulhas leakage with northern summer insolation suggests a teleconnection to the monsoon system. The onset of increased Agulhas leakage during late glacial conditions took place when glacial ice volume was maximal, suggesting a crucial role for Agulhas leakage in glacial terminations, timing of interhemispheric climate change⁷ and the resulting resumption of the Atlantic meridional overturning circulation⁸.

The magnitude and intensity of the transport flux of heat and salt between the Indian and Atlantic oceans play a decisive role in global ocean circulation and the Atlantic meridional overturning circulation^{1–6,8–11}. Surface and intermediate water transport from the Indian into the Atlantic Ocean takes place through the detachment of rings, eddies, and filaments from the Agulhas current and retroflection^{2,10}. Whereas the modern hydrographic variability of the Agulhas system has been studied extensively^{1–3,8–10}, the palaeo-oceanography on orbital timescales (10^4 – 10^5 yr) is virtually uninvestigated, although some records have appeared^{12,13}. Here we report new findings on the unique planktic foraminiferal signature of a modern Agulhas ring. We use this information as a modern analogue to disclose the late Pleistocene evolution of Agulhas transport into the South Atlantic using fossil planktic foraminifera.

Surface waters in the Cape basin may be derived from three different regions^{2,9,10}: (1) the Indian Ocean's Agulhas current, (2) the South Atlantic current, north of the Sub-Tropical Convergence (STC), and (3) the subantarctic surface waters south of the STC. The MARE expeditions¹⁴ examined a ~300-km-wide Agulhas ring called 'Astrid' (Fig. 1a), spawned from the Agulhas retroflection in January 2000 (ref. 15). Its faunal content was sampled by using depth-stratified plankton towing in the centre, near the boundary and the outside of the ring in March 2000, July/August 2000 and February 2001 (see Supplementary Information). These modern observations show that the ring is characterized by typical tropical–subtropical species^{16,17} (Table 1). The most important are *Globigerinoides ruber*, *Globigerinoides sacculifer*, *Globigerinella siphonifera*, *Globigerinita glutinata*, *Orbulina universa*, *Globorotalia menardii* and *Globoquadrina hexagona* (Table 1, Methods). The modern distribution of living and seafloor fossil planktic foraminifera in the Indian and Atlantic oceans is discussed extensively elsewhere^{16–18}. Identification of Agulhas fauna in ring Astrid agrees well with previous findings that the Agulhas current transports these species into the Cape basin. Faunal signatures can now be assigned to each of the Cape basin surface water masses. The Agulhas leakage water is traced by the Agulhas leakage fauna (ALF, see Methods), whereas water masses of the subpolar and polar zones are characterized by *Turborotalita quinqueloba* and *N. pachyderma* (sin.)^{16–18} (Fig. 1b). The species *Globorotalia inflata* and *Neogloboquadrina pachyderma* (dex.) are mostly found in waters of the southern subtropical to subantarctic zones, while subtropical to transitional waters mainly north of the STC (that is, north of ~42° S) are reflected by *Globorotalia truncatulinoides*. Consequently, we will use the ratio of *G. truncatulinoides*/(*N. pachyderma* (dex.) + *G. inflata* + *G. truncatulinoides*) as a proxy for the distance of the STC relative to our sediment core sites in the southern Cape Basin.

The late Pleistocene history of Agulhas leakage is reconstructed using planktic foraminiferal assemblage data from two sediment cores located precisely beneath the 'Agulhas ring corridor' in the

southern Cape basin. The cores GeoB-3603-2 and MD96-2081 were spliced (see Methods) to form the 550-kyr-old Cape Basin record (CBR). Characteristic changes and events in the down-core oxygen isotope ratio of the benthic foraminifer *Cibicides wuellerstorfi*

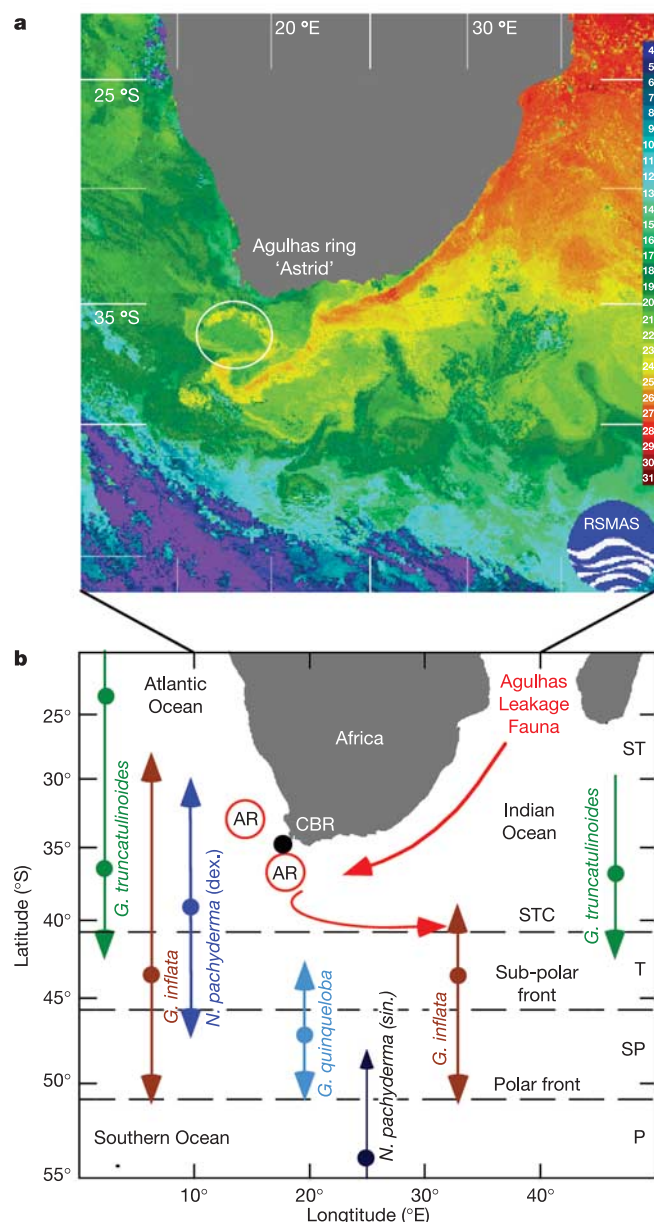


Figure 1 SST and latitudinal distribution of some planktic foraminifer species. **a**, Satellite image showing the SST of the Agulhas region in March 2000. The image clearly shows the extreme temperature contrast between the Agulhas current east of Africa and the southeast Atlantic. The circle indicates the position of Agulhas ring 'Astrid'¹⁵, surrounded by a warm water filament, two months after spawning from the retroflection. The ring's faunal content was sampled by using depth-stratified plankton towing in the centre, the boundary and the outside of the ring in March 2000, July/August 2000 and February 2001 (see Supplementary Information). Image made available by the Rosenstiel School of Marine and Atmospheric Sciences, Miami, USA (<http://www.rsmas.miami.edu/environment/imagery/>). **b**, Schematic representation of the latitudinal distribution of some species of planktic foraminifera, as indicated by the arrows and position of frontal zones. The dots reflect the approximate position of the modern maximum relative abundance of the species^{16–18}. Red arrows indicate transport direction of tropical and subtropical species, by the Agulhas current and retroflection. Circles labelled 'AR' show the pathway of Agulhas rings into the south Atlantic Ocean¹⁰. P, polar zone; SP, subpolar zone; T, transitional zone; ST, subtropical zone.

($\delta^{18}\text{O}_{\text{benthic}}$) served as control points for the age model tuned to the standard orbital chronostratigraphy¹⁹ (see Supplementary Information) (Fig. 2a). A foraminifera-independent proxy for sea surface temperature²⁰ (SST) is available for the upper part of the CBR (core GeoB-3603-2) from ketone unsaturation ratio (U_{37}^{K}) measurements (Figs 2b, 3e–i). Preservation of the highly abundant planktic foraminifera in the CBR is very good (see Supplementary Information) and species diversity is high (thirty species), reflecting the convergence of several major water masses. Records of the ALF and species vital to the understanding of past variability in Agulhas leakage are shown in Fig. 2c–h.

The (U_{37}^{K}) SST and ALF proxy records show strong glacial–interglacial variation. Enhanced Indian–Atlantic transport is observed during interglacial stages (Figs 2, 3). During glacial stages the relative contribution of ALF is low, indicating that Indian–Atlantic communication was reduced or may have temporarily ceased during MIS-12. The initiation of the late glacial SST rise precedes maximum ice volume, as indicated by the $\delta^{18}\text{O}_{\text{benthic}}$ proxy (Figs 2a, b and 3e–h), while late glacial initiation of enhanced Agulhas leakage starts when glacial ice volume is maximal (Fig. 3e–i). The ALF and SST rapidly increase during the first half of the glacial terminations V to I (Fig. 3e–i) and were maximal during the second half. During interglacial MIS-11, Agulhas leakage was persistent over a period of ~60,000 yr—that is, from 430–370 kyr ago.

Simultaneous occurrence of subantarctic species and increased Agulhas leakage during MIS-9 and MIS-8 (Fig. 2g,h) may point to a narrowing of the leakage passage. The northward shift of frontal zones, as reflected in the STC proxy (Fig. 3c) and by the higher relative abundance of subantarctic–Antarctic species (Fig. 2g,h), must have resulted in increased northward transport of cold wedges of subantarctic surface water during the formation of rings at the Agulhas retroflection^{10,12}. Narrowing of the Agulhas corridor may alternatively explain the anomalous occurrences of subantarctic and Antarctic species in the southeast Atlantic^{21,22}. Leakage reaches a maximum again during Termination II (Figs 2c and 3b,f). From this maximum, the leakage declined through the later part of MIS-5 and into the period MIS 4 to MIS 2. Establishment of modern values occurred during MIS 2 and Termination I into the Holocene. The

Table 1 Living planktic foraminiferal assemblages off South Africa

Species name	PCA-1	PCA-2	PCA-3	PCA-4	PCA-5
<i>Pulleniatina obliquiloculata</i> *	0.98	0.00	0.09	0.06	0.04
<i>Globigerinita glutinata</i> *	0.96	0.09	0.07	0.17	0.01
<i>Hastigerina pelagica</i> *	0.90	0.12	0.01	0.03	–0.03
<i>Globorotalia menardii</i> *	0.86	0.25	0.06	–0.09	0.10
<i>Globigerinella calida</i>	0.70	–0.05	0.68	0.00	0.07
<i>Globigerinoides sacculifer</i> *	–0.05	0.96	–0.05	0.08	0.06
<i>Globigerinella siphonifera</i> *	0.21	0.92	0.12	0.01	0.08
<i>Globigerinoides ruber</i> *	0.33	0.88	–0.07	0.20	0.01
<i>Orbulina universa</i> *	0.00	0.86	0.01	0.28	0.14
<i>Neoglobobulimina dutertrei</i> †	0.09	0.54	0.04	0.75	–0.11
<i>Globorotalia hirsuta</i>	–0.14	–0.04	0.93	–0.08	–0.01
<i>Globorotalia truncatulinoides</i>	0.29	–0.05	0.91	–0.05	0.10
<i>Globorotalia inflata</i>	0.09	0.25	0.77	0.50	0.06
<i>Globigerina bulloides</i>	0.29	0.29	0.15	0.80	–0.01
<i>Neoglobobulimina pachyderma (dex.)</i>	–0.13	–0.02	–0.11	0.71	0.22
<i>Neoglobobulimina pachyderma (sin.)</i>	–0.11	–0.19	–0.18	0.18	0.67
<i>Globobulimina hexagona</i> *	–0.04	–0.13	–0.15	–0.02	–0.60
<i>Globorotalia scitula</i> *	–0.11	–0.15	–0.07	0.02	–0.51
Cumulative percentage of variance explained.	24	45	61	73	80

Numbers indicate the species' loading on Varimax-rotated principal components. Loadings greater than 0.50 or less than –0.50 are printed in boldface. Species marked with an asterisk are included in the Agulhas leakage fauna.

Interpretation of principal component assemblages. PCA-1, Agulhas leakage species with depth habitat 0–300 m; PCA-2, Agulhas leakage species with depth habitat 0–200 m (†*N. dutertrei* not included in reconstruction); PCA-3, Species indicative for the subtropical to transitional zones with depth habitat 100–500 m; PCA-4, Species indicative for the transitional-subantarctic zones with depth habitat 0–200 m; PCA-5, Positive loading, that is, *N. pachyderma* (sin.), indicative for subantarctic to Antarctic zone. Negative loading indicates Agulhas leakage species with depth habitat 300–800 m.

faunal and SST record of the southern Benguela system²² exhibit a similar pattern of late glacial increase of tropical species and SST over the last five glacial terminations, confirming the uptake of Agulhas water into the southeast Atlantic Ocean.

Variation in the ALF and STC proxy show a strong 100-kyr periodicity (Fig. 3b, c), indicating that the most dominant climate signal of Agulhas leakage variability is related to the periodic glacial–interglacial changes. The expansion of the Antarctic ice sheet, changes in the wind field², increased sea ice cover²³ and a northward position of the STC and frontal zones^{13,24}, may have contributed to the reduction of Agulhas leakage during glacial periods. A northward shift of the wind pattern results in a reduction of Agulhas leakage². As the STC, reflecting the zero wind-curl line, approaches the tip of the continent, the southward-penetrating Agulhas effectively seals off the corridor through which Indian Ocean water leaks into the south Atlantic. Figure 3c shows that the STC reached its most northward position during mid-glacial conditions and not under late glacial conditions when continental ice volume was largest, effectively reducing the leakage into the South Atlantic.

The main orbital parameters affecting solar insolation and its distribution may have controlled the position of the STC and variations in Agulhas leakage. The dominant glacial–interglacial 100-kyr cycle is clearly present in most proxy records of the Cape basin (Figs 2b–f, 3a–c) and is most pronounced in the STC proxy (see Supplementary Information), showing most northward STC positions at minimum eccentricity (Fig. 3a,c). This implies that the amplitude modulation of precession by eccentricity (Fig. 3a) affects not only ice volume and tropical SST²⁰, but also the pattern of the wind field, the associated position of the STC and the variability of Agulhas leakage. Superimposed on the large-scale glacial–interglacial 100-kyr variability, the Agulhas leakage signal includes moderate tilt (41 kyr) and precessional (23 and 19 kyr) power (see Supplementary Information), implying both a ‘low’ and mid-high latitude forcing of the leakage. The relationship of the ALF to tilt is obviously related to the position of the frontal zones such as the STC and the associated shift in SSTs²⁵. Comparison of the ALF with the solar insolation curves revealed that maxima in northern summer insolation and minima in precession are linked to maxima in ALF (Fig. 3a, b). We suggest that the increased strength of the monsoon

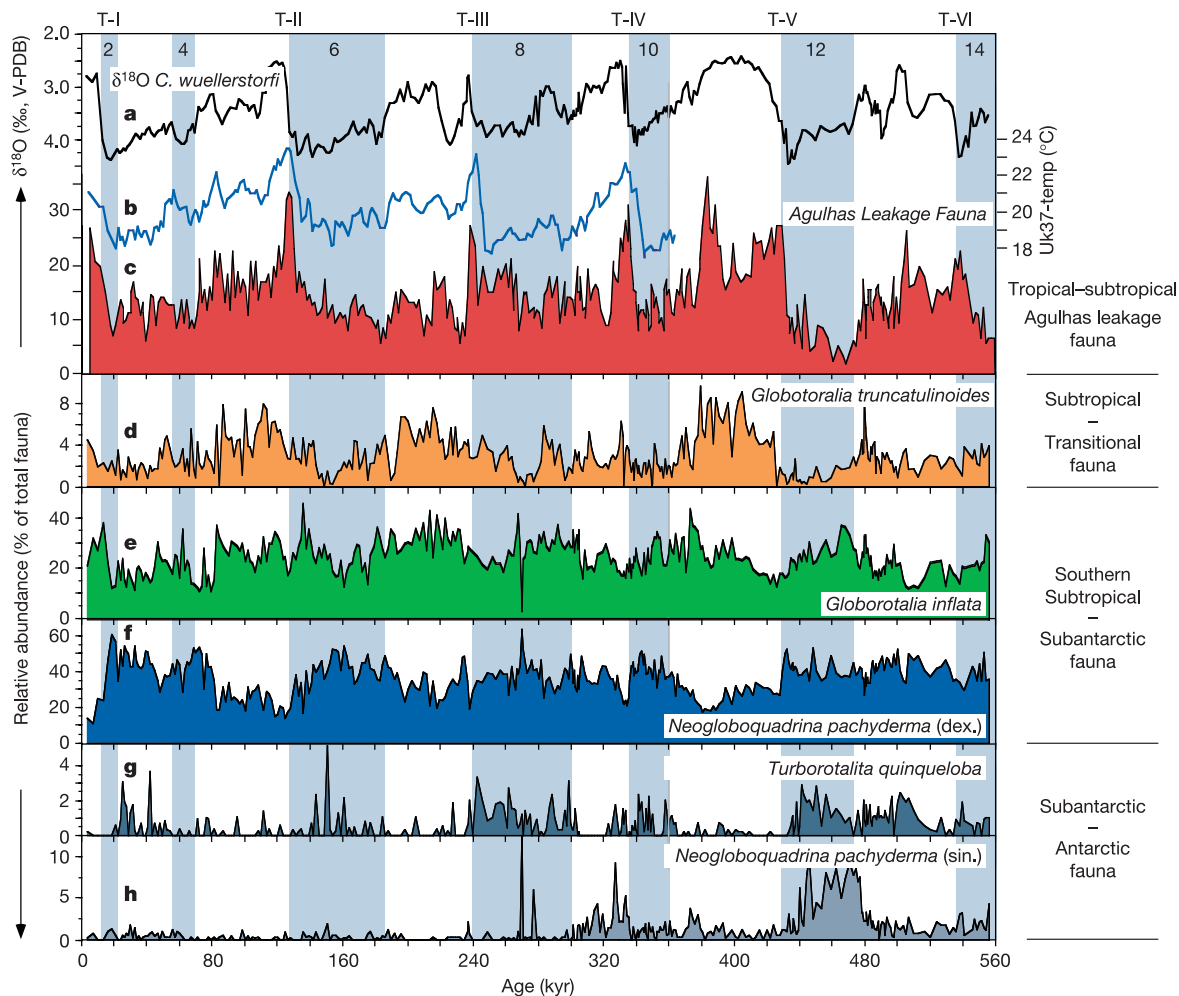


Figure 2 Late Pleistocene proxy records of the Cape Basin record. **a**, The oxygen isotope composition of the benthic foraminifer *Cibicides wuellerstorfi*, largely indicating global ice volume (V-PDB, Vienna Pee-Dee Belemnite standard). Even numbers at top refer to glacial Marine Isotopic Stages, T-I to T-VI refer to the major glacial terminations. **b**, The $(U_{37}^{K'})$ SST proxy for the upper part of the CBR (core GeoB-3603-2). **c**, The relative abundance of the Agulhas leakage fauna as a measure of Indian Ocean advection into the Atlantic.

d, The relative abundance of *Globorotalia truncatulinoides*, indicative of subtropical-transitional waters. **e**, The relative abundance of *Globorotalia inflata* and **f**, *Neogloboquadrina pachyderma* (dex.) indicative of southern subtropical to subantarctic waters. **g**, The relative abundance of *Turborotalita quinqueloba*, reflecting subantarctic waters. **h**, The relative abundance of *Neogloboquadrina pachyderma* (sin.), indicative of Antarctic waters^{16–18}.

system promoted the equatorial transport of warm tropical waters feeding the Agulhas current and increased the Agulhas leakage. Based on an analysis of satellite altimetry data a relationship between the strength and variability of the Indian Ocean equatorial winds and the formation of rings from the Agulhas retroflexion was recently proposed²⁶. The signal propagates via a succession of Kelvin and Rossby waves that seems eventually to trigger the formation of rings. This implies that a stronger monsoon system results in an increase of the strength of this oceanic teleconnection between the equatorial and the southwest Indian Ocean, enhancing Agulhas ring formation and leakage.

Strongly guided by the African slope and Agulhas plateau, it is unlikely the Agulhas current would have deviated strongly from its present position during low sea level. A sediment record from beneath the Agulhas current to the south of and east of South Africa indicates that the glacial Agulhas current continued to flow and did not shift much laterally over the last 150,000 yr (ref. 27). The

reduced glacial leakage is therefore probably caused by a further eastward retroflexion of the Agulhas current, as has also been observed recently²⁸. A northward shift of the STC and synchronous eastward forcing of the retroflexion may have led to increased recirculation of tropical Agulhas water within the southern Indian Ocean. Such a mechanism would effectively transmit tropical signals to the southern Indian and Pacific oceans and might explain the increased glacial climate variability, with 'a northern hemispheric signature', in regions of the STC in the southern Indian and Pacific oceans²⁹.

The observation that modern Agulhas rings are characterized by an Indian Ocean faunal signature allows us to assess the variability of Indian–Atlantic water communication. The reconstructed history of Agulhas leakage provides important new data and insights into the processes and responses of climate-driven ocean circulation and places constraints on possible mechanisms of interhemispheric climate timing⁷. The late glacial rise in SST before the start of the

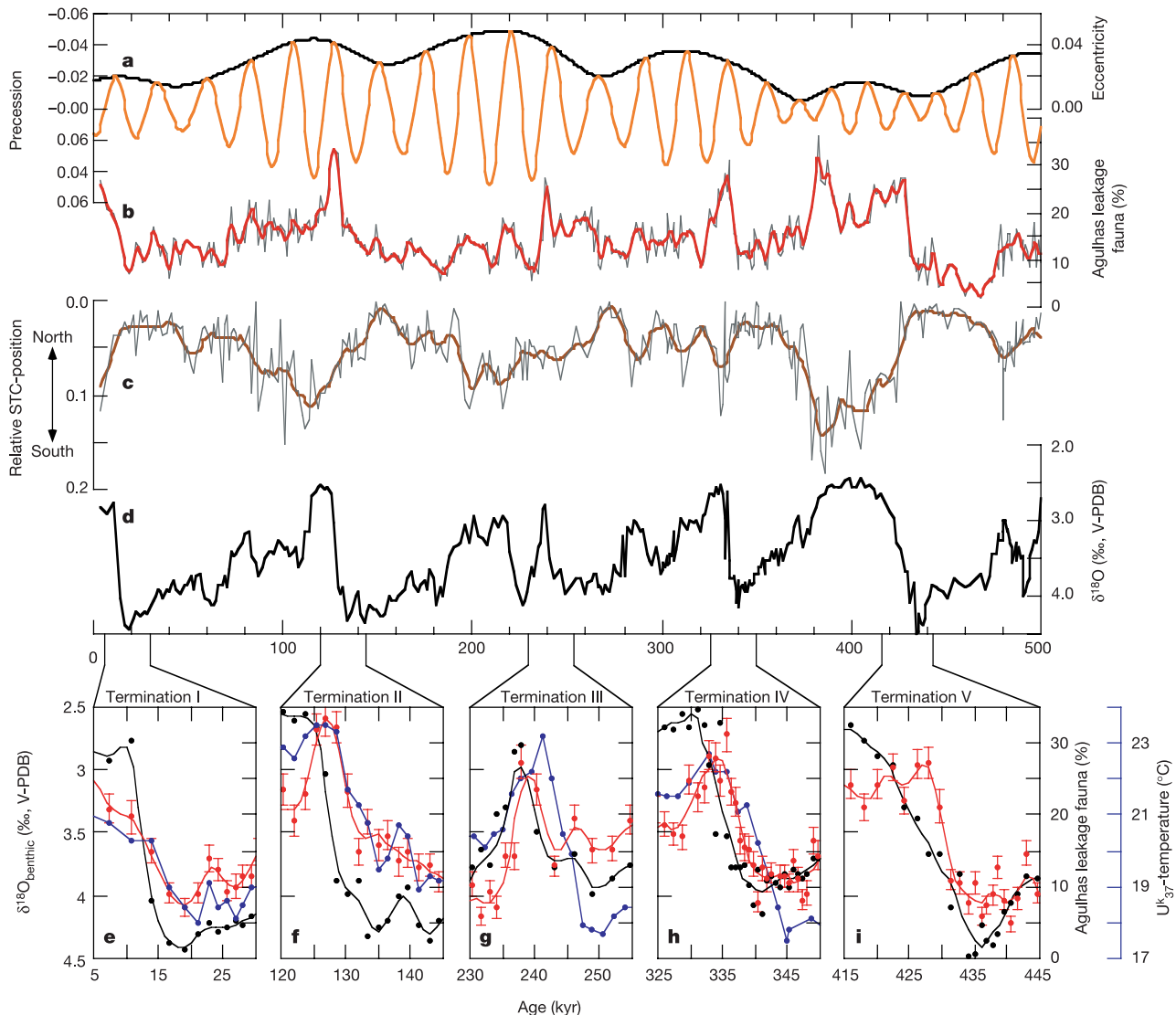


Figure 3 Late Pleistocene orbital variability and Cape basin proxy records during glacial terminations. **a**, Orbital eccentricity (black line) and precession (note inverted axis). **b**, The Agulhas leakage fauna. Periods of maximum Agulhas leakage appear to be related to minima in precession (maxima in northern summer insolation). **c**, The ratio of *G. truncatulinoides*/(*N. pachyderma* dex. + *G. inflata* + *G. truncatulinoides*) as a proxy for the distance of the STC relative to the position of the Cape basin record. High values

indicate a southward position for the STC (note inverted axis). **d**, The $\delta^{18}\text{O}$ of *Cibicides wuellerstorfi*. **e–i**, The Cape basin proxy records of Agulhas leakage (red line), global ice volume (black line) and SST (blue line) during the major late Pleistocene glacial terminations I–V. Smoothed curves ($\delta^{18}\text{O}_{\text{benthic}}$ and ALF) represent seven-point smoothing of the 1-kyr sampled data. Original data indicated with dots. Error bars on the Agulhas leakage fauna specify one standard error of the mean.

major deglaciations and the onset of enhanced leakage when glacial ice volume is maximal underscore the potential role of Agulhas leakage in the resumption of Atlantic thermohaline circulation and North Atlantic Deep Water (NADW) formation^{1–5,10}. This is confirmed by carbon and oxygen isotope records from cores located in the south Atlantic sector of the Southern Ocean³⁰. Here, strong input of NADW into the Southern Ocean is observed during periods of maximum leakage. Our results support the conclusion of recent modelling results⁶—that abrupt resumption of the interglacial mode of the thermohaline circulation after glacial conditions is probably triggered by increased mass transport of Indian Ocean water into the Atlantic Ocean via the Agulhas system.

The faunal record also implies that the onset of a southward-moving STC, after reaching its most northward position during mid-glacial conditions, started well before the terminations and maxima in the leakage. Hence, the Indian–Atlantic connection via the Agulhas leakage should be considered an important marine amplifier for the 100-kyr cycle that dominates glacial–interglacial changes during the late Pleistocene. The flow of Indian Ocean water into the Atlantic is therefore indicative of changes in the global ocean–climate system and a precursor to re-establishment and maintenance of full interglacial conditions. □

Methods

The Agulhas leakage fauna

The Agulhas leakage fauna (ALF) includes all species loading greater than 0.85 on the first and second principal component axis (PCA) and species loading less than −0.50 on the fifth PCA (Table 1). The ALF thus includes the species: *Pulleniatina obliquiloculata*, *Globigerinita glutinata*, *Hastigerina pelagica*, *Globorotalia menardii*, *Globigerinoides sacculifer*, *Globigerinella siphonifera*, *Globigerinoides ruber*, *Orbulina universa*, *Globoquadrina hexagona* and *Globorotalia scitula*. This grouping appeared most consistent with the plankton tow observations when each station is studied one by one with respect to their position on the Agulhas ring studied (see also Supplementary Information). The ALF proxy used in this study is calculated as the sum of the relative abundance of the species listed above.

The Cape basin record

Two sediment cores, GeoB-3603-2 and MD96-2081 were used to form one time series. Both cores are located in very close proximity to each other, at similar depths within the southern Cape basin. Core GeoB-3603-2 was retrieved from 35° 08' S to 17° 33' E (2,840 m depth), while core IMAGES-II MD96-2081 is located at 35° 35' S to 17° 41' E (3,164 m depth). The cores are positioned on the upper slope, parallel to but south of Cape Canyon. Extensive analysis of the oxygen isotope data from both cores reveals that they overlap between 365 and 238 kyr. The joining point was taken at the peak of MIS 7.5. Samples in the overlapping section were removed to produce a single record which covers a continuous period of ~550 kyr. The upper portion of the spliced record uses samples from core GeoB-3603-2 and the lower section is core MD96-2081. Eighty-six of the samples from the bottom of core GeoB-3603-2 (708–1,133 cm below sea floor) are not included in the Cape basin record. The record described in this work relates to this spliced core and is referred to as the CBR. All characteristic glacial and interglacial stages of the Cape basin benthic $\delta^{18}\text{O}$ record were identified down to MIS 14 and tuned to the SPECMAP record, yielding an approximate age for the base of the record of ~550 kyr. Using this method MIS 5.5, 9 and 11 are isotopically the lightest. The glacial periods MIS 2, 6 and 10 have similarly high $\delta^{18}\text{O}$ values, whereas MIS 12 shows highest values during the late Pleistocene, indicating the most extreme glacial period of the time studied. More information on both cores can be found in the Supplementary Information.

Sampling resolution and faunal counts

Both cores were sampled every 5 cm, which produces an average temporal interval of ~1.5 kyr between two successive samples. Planktic foraminiferal counts were carried out on the >150- μm size fraction. To facilitate examination and counting, the fraction was divided using a microsplitter until approximately 400 specimens remained for counting. All faunal counts were based on at least 300 specimens per sample.

Spectral analysis

A Blackman–Tukey spectral cross-correlation between Northern Hemisphere summer insolation (21 June at 60° N) and (1) the Agulhas leakage fauna, and (2) the ratio of *Globorotalia truncatulinoides*/(*Neogloboquadrina pachyderma* (dex.) + *Globorotalia inflata* + *Globorotalia truncatulinoides*), was performed using the AnalyseSeries software version 1.2 of ref. 31. The time interval between 4 and 500 kyr was analysed, using the 'compromise' option between resolution and confidence. This resulted in a number of lags of 149, or 30% of the series. The faunal abundance and oxygen isotope data were re-sampled every 1 kyr using the linear function integration method. The data were linearly detrended and normalized to unit variance before analysis. Bandwidth, 0.010.

Received 12 February; accepted 25 June 2004; doi:10.1038/nature02785.

- Gordon, A. L. Communication between oceans. *Nature* **382**, 399–400 (1996).
- de Ruijter, W. P. M. *et al.* Indian–Atlantic interocean exchange: Dynamics, estimation and impact. *J. Geophys. Res.* **104**, 20885–20910 (1999).
- Gordon, A. L. The branniest retroflection. *Nature* **421**, 904–905 (2003).
- Weijer, W., de Ruijter, W. P. M. & Dijkstra, H. A. Stability of the Atlantic overturning circulation: Competition between Bering Strait freshwater flux and Agulhas heat and salt sources. *J. Phys. Oceanogr.* **31**, 2385–2402 (2001).
- Weijer, W., De Ruijter, W. P. M., Sterl, A. & Drijfhout, S. S. Response of the Atlantic overturning circulation to South Atlantic sources of buoyancy. *Glob. Planet. Change* **34**, 293–311 (2002).
- Knorr, G. & Lohmann, G. Southern Ocean origin for the resumption of Atlantic thermohaline circulation during deglaciation. *Nature* **424**, 532–536 (2003).
- Wunsch, C. Greenland–Antarctic phase relations and millennial time-scale climate fluctuations in the Greenland ice-cores. *Quat. Sci. Rev.* **22**, 1631–1646 (2003).
- Gordon, A. L., Weiss, R. F., Smethie, W. M. Jr & Warner, M. J. Thermocline and intermediate water communication between the South Atlantic and Indian oceans. *J. Geophys. Res.* **97**, 7223–7240 (1992).
- Garzoli, S. L. & Gordon, A. L. Origins and variability of the Benguela Current. *J. Geophys. Res.* **101**, 897–906 (1996).
- Lutjeharms, J. R. E. in *The South Atlantic: Present and Past Circulation* (eds Wefer, G., Berger, W. H., Siedler, G. & Webb, D. J.) 125–162 (Springer, Berlin/Heidelberg, 1996).
- Macdonald, A. M. & Wunsch, C. An estimate of global ocean circulation and heat fluxes. *Nature* **382**, 436–439 (1996).
- Rau, A. J. *et al.* A 450-kyr record of hydrological conditions on the western Agulhas Bank Slope, south of Africa. *Mar. Geol.* **180**, 183–201 (2002).
- Flores, J.-A., Gersonde, R. & Sierro, F. J. Pleistocene fluctuations in the Agulhas Current Retroflection based on the calcareous plankton record. *Mar. Micropaleontol.* **37**, 1–22 (1999).
- Lutjeharms, J. R. E. *et al.* MARE and ACSEX: new research programmes on the Agulhas Current System. *S. Afr. J. Sci.* **96** (2000).
- van Aken, H. M. *et al.* Observations of a young Agulhas ring Astrid, during MARE in March 2000. *Deep-Sea Res. II* **50**, 167–195 (2003).
- Bé, A. W. H. & Tolderlund, D. S. in *Micropaleontology of Oceans* (eds Funnell, B. M. & Riedel, W. R.) 105–149 (Cambridge Univ. Press, London, 1971).
- Bé, A. W. H. & Hutson, W. H. Ecology of planktonic foraminifera and biogeographic patterns of life and fossil assemblages in the Indian ocean. *Micropaleontology* **23**, 369–414 (1977).
- Niebler, H.-S. & Gersonde, R. A planktic foraminiferal transfer function for the southern South Atlantic Ocean. *Mar. Micropaleontol.* **34**, 213–234 (1998).
- Imbrie, J. *et al.* in *Milankovitch and Climate I* (eds Berger, A. L., Imbrie, J., Hays, J. D., Kukla, J. & Salzman, J.) 269–305 (Reidel, Hingham, Massachusetts, 1984).
- Schneider, R. R., Müller, P. J. & Acheson, R. in *Reconstructing Ocean History: A Window into the Future* (eds Abrantes, F. & Mix, A.) 35–55 (Kluwer Academic/Plenum, New York, 1999).
- Ufkes, E., Jansen, J. H. F. & Schneider, R. R. Anomalous occurrences of *Neogloboquadrina pachyderma* (left) in a 420-kyr upwelling record from Walvis Ridge (SE Atlantic). *Mar. Micropaleontol.* **40**, 23–42 (2000).
- Chen, M.-T. *et al.* Late Quaternary sea-surface temperature variations in the southeast Atlantic: a planktic foraminiferal faunal record of the past 600,000 yr (IMAGES II MD962085). *Mar. Geol.* **180**, 163–181 (2002).
- Kunz-Pirrung, M., Gersonde, R. & Hodell, D. A. Mid-Brunhes century-scale diatom sea surface temperature and sea ice records from the Atlantic sector of the Southern Ocean (ODP Leg 177, Sites 1093, 1094 and core PS2089–2). *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **182**, 305–328 (2002).
- Howard, W. R. & Prell, W. L. Late Quaternary surface circulation of the southern Indian Ocean and its relationship to orbital variations. *Paleoceanography* **7**, 79–117 (1992).
- Sachs, J. P., Anderson, R. F. & Lehman, S. J. Glacial surface temperatures of the southeast Atlantic Ocean. *Science* **293**, 2077–2079 (2001).
- Schouten, M. W., de Ruijter, W. P. M., van Leeuwen, P. J. & Dijkstra, H. A. An oceanic teleconnection between the equatorial and southern Indian Ocean. *Geophys. Res. Lett.* **29**(16), doi:10.1029/2001GL014542 (2002).
- Winter, A. & Martin, K. Late Quaternary history of the Agulhas current. *Paleoceanography* **5**, 479–486 (1990).
- de Ruijter, W. P. M. *et al.* Eddies and dipoles around South Madagascar: formation, pathways and large-scale impact. *Deep-Sea Res. I* **51**, 383–400 (2004).
- Pahnke, K., Zahn, R., Elderfield, H. & Schulz, M. 340,000-year centennial-scale marine record of Southern Hemisphere climatic oscillation. *Science* **301**, 948–952 (2003).
- Hodell, D. A., Charles, C. & Ninnemann, U. S. Comparison of interglacial stages in the South Atlantic sector of the southern ocean for the past 450 kyr: implications for Marine Isotope Stage (MIS) 11. *Glob. Planet. Change* **24**, 7–26 (2000).
- Paillard, D., Labeyrie, L. & Yiou, P. Macintosh program performs time-series analysis. *Eos Trans. AGU* **77**, 379 (1996).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements The MARE/CLIVARNET project was sponsored by two NWO grants (bilateral cooperation programme NWO-DFG). R.A. acknowledges funding from a NERC grant. R.S. is grateful to the German Science Foundation (DFG) and the IMAGES programme for sponsoring core retrieval and analytical work. We thank M. Segl and P. Mueller for measuring isotopes and alkenones, respectively. We thank the captain and crew of R/V *Pelagia* and R/V *Agulhas* as well as the Royal NIOZ technicians. We thank all MARE team members for discussion. This is a publication of the Netherlands Research School of Sedimentary Geology (NSG).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to F.P. (frank.peeters@falw.vu.nl).

The avian nature of the brain and inner ear of *Archaeopteryx*

Patricio Domínguez Alonso¹, Angela C. Milner², Richard A. Ketcham³, M. John Cookson⁵ & Timothy B. Rowe⁴

¹Departamento de Paleontología, Universidad Complutense de Madrid, Ciudad Universitaria, 28040 Madrid, Spain

²Department of Palaeontology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK

³High-resolution X-ray CT Facility, ⁴Department of Geological Sciences, 1 University Station C1110, University of Texas at Austin, Austin, Texas 78712, USA

⁵University of Hertfordshire, Hatfield, Herts AL10 9AB, UK

Archaeopteryx, the earliest known flying bird (avian) from the Late Jurassic period, exhibits many shared primitive characters with more basal coelurosaurian dinosaurs (the clade including all theropods more bird-like than *Allosaurus*)¹, such as teeth, a long

bony tail and pinnate feathers². However, *Archaeopteryx* possessed asymmetrical flight feathers on its wings and tail, together with a wing feather arrangement shared with modern birds. This suggests some degree of powered flight capability³ but, until now, little was understood about the extent to which its brain and special senses were adapted for flight. We investigated this problem by computed tomography scanning and three-dimensional reconstruction of the braincase of the London specimen of *Archaeopteryx*. Here we show the reconstruction of the braincase from which we derived endocasts of the brain and inner ear. These suggest that *Archaeopteryx* closely resembled modern birds in the dominance of the sense of vision and in the possession of expanded auditory and spatial sensory perception in the ear. We conclude that *Archaeopteryx* had acquired the derived neurological and structural adaptations necessary for flight. An enlarged forebrain suggests that it had also developed enhanced somatosensory integration with these special senses demanded by a lifestyle involving flying ability⁴.

The London specimen of *Archaeopteryx* BMNH 37001 is the only one of the seven known in which high resolution computed

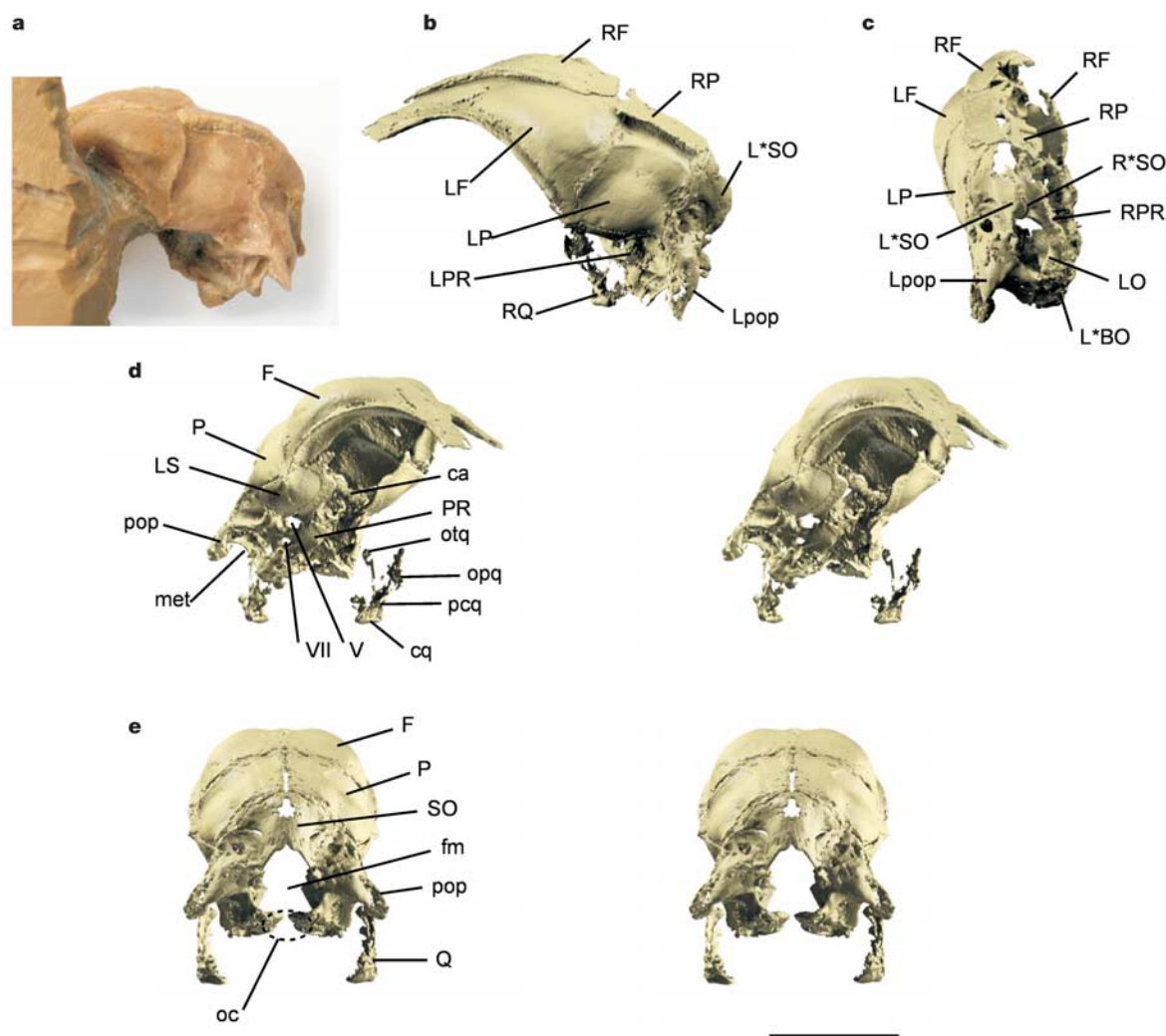


Figure 1 Braincase of the holotype of *Archaeopteryx lithographica* (BMNH 37001). Scale bar, 10 mm. **a**, Original prepared specimen in left lateral view⁵. **b, c**, 3D reconstructions of braincase based on X-ray CT data; left lateral view (**b**) and posterior view (**c**) showing right side collapse. **d, e**, Stereopair of restored braincase in oblique right-anterior view (**d**) and posterior view (**e**). Elements from the left side are reversed except for the right quadrate. L and R indicate left and right. Abbreviations: *BO?, ?basioccipital fragment; ca, crista

arcuata; cq, mandibular condyle of quadrate; F, frontal; fm, foramen magnum; LS, laterosphenoid; met, metotic; O, opisthotic; oc, occiput; opq, orbital process of quadrate; otq, otic wing of quadrate; P, parietal; pcq, pterygoid condyle of quadrate; pop, paroccipital process of opisthotic; PR, prootic; Q, right quadrate; *SO, supraoccipital fragment; V, trigeminal nerve foramen; VII, facial nerve foramen. Original imagery available at < <http://www.DigiMorph.org> > .

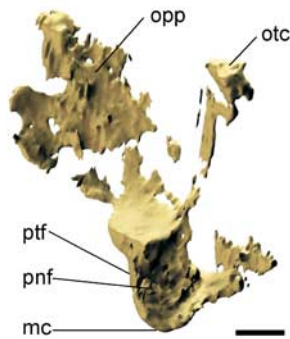


Figure 2 Right quadrate of BMNH 37001 in medial view. Scale bar, 1 mm. Abbreviations: ptf, pterygoid facet; pnf, pneumatic foramen; mc, mandibular condyle; opp, optic process; otc, otic capitulum. See also Supplementary Information for clear views of the pneumatic foramen.

tomography (CT) scanning⁵ could be attempted because the braincase (Fig. 1a), including a partially exposed endocast (Fig. 1b), had been removed from the edge of the main specimen slab and prepared⁶ in left lateral and posterior view to facilitate its description^{7,8}. The right side of the braincase, still encased in matrix, is collapsed (Fig. 1c) and the floor is absent.

Our braincase reconstruction (Fig. 1d, e) shows that the frontals have an anterior articular facet for the nasals, and that the parietals, frontals and laterosphenoids meet in a blunt postorbital process. At this point, their margins are not clearly definable suggesting possible fusion. A crescentic fontanelle is present along the upper margin of the right crista arcuata. This is a vascular opening allowing the dural venous sinus (caudal petrosal sinus in avian literature) to drain out of the skull. Our reconstruction confirms that the previous identification of the opening on the occiput as the foramen for the external occipital vein^{7,8} is correct. The foramen magnum is broad (40% of the total width of the brain) and triangular, in contrast to the smaller rounded shape reconstructed previously⁷.

The right quadrate (Figs 1d and 2), hitherto unidentified and buried in the matrix, is almost in its original anatomical position. A similar quadrate was reported in the Munich specimen^{9,10} and the morphology confirms that an isolated element previously reported in BMNH 37001 as a right quadrate⁷ was incorrect, as suspected¹⁰; we suggest that it might be a sphenoid fragment. The form of the pterygoid articulation, at the limit of the CT resolution, appears to be rounded and therefore a condyle rather than a facet, which suggests a possibly mobile quadrate-pterygoid. The mesial face is perforated by a pneumatic foramen, positioned just dorsal to the mandibular condyle and just posterior to the pterygoid facet, that represents a previously unreported pneumatic cavity in *Archaeopteryx*. Witmer¹¹ considered quadrate pneumaticity to be a synapomorphy of at least carinate birds and the position of this foramen in *Ichthyornis* to be primitive for birds. Its presence in *Archaeopteryx* suggests it may be a synapomorphy at the avialan level. However, the distribution of quadrate pneumaticity in non-avian theropods is complex; it is present in a few groups including some maniraptorans (coelurosaurs with a bird-like flexible wrist) but is lacking in dromaeosaurids¹².

We have reconstructed most of the brain in three dimensions (3D) (Fig. 3). Previous models have necessarily relied on the surface details visible in BMNH 37001 (refs 13, 14) and other *Archaeopteryx* specimens, principally the Eichstätt and Munich individuals⁹. Pterosaurs, small coelurosaurs and birds have thin meninges and small venous sinuses so that the endocast of the braincase reflects precisely the brain anatomy and can be considered as an accurate index of the brain volume, mass and surface morphology^{15–17}.

The crescentic olfactory bulbs join the cerebral lobes by means of

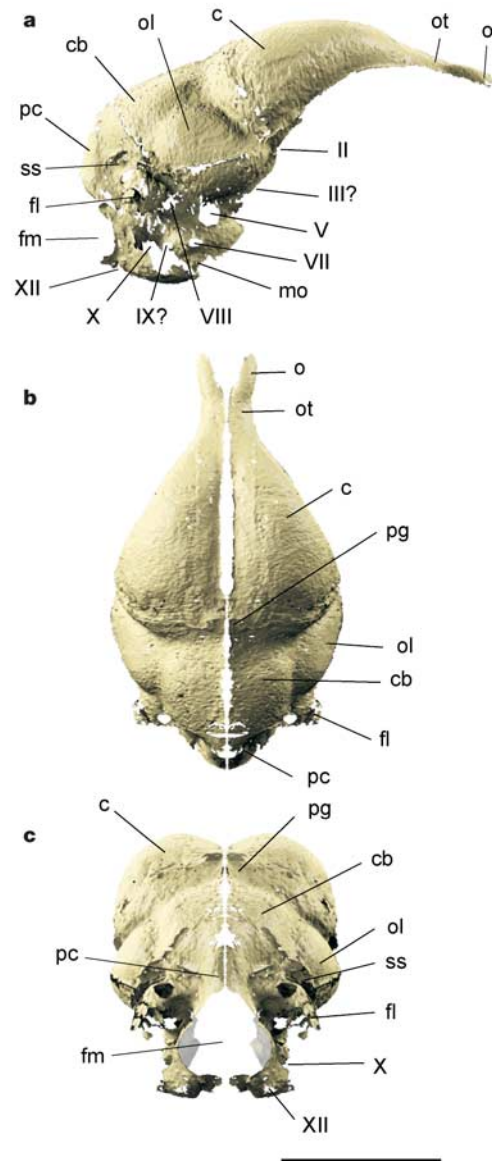


Figure 3 Restored endocast of the brain of BMNH 37001 rendered as a shell. Elements from the left side are reversed. Scale bar, 10 mm. **a**, Right lateral view. **b**, Dorsal view. **c**, Posterior view. Abbreviations: c, cerebrum (telencephalon); cb, cerebellum; fl, floccular lobe of the cerebellum; fm, foramen magnum; mo, medulla oblongata (rhombencephalon); o, olfactory bulb; ol, optic lobes (metencephalic tectum); ot, olfactory tract; pc, cerebellar prominence; pg, pineal gland (epiphysis); ss, semicircular sulcus. Cranial nerves: II, optic; III?, possible oculomotor; V, trigeminal; VII, facial; VIII?, auditory (or endolymphatic duct, see Fig. 5); IX?, glossopharyngeal; X, vagus; XII, hypoglossal. Original imagery available at < <http://www.DigiMorph.org> > .

thick olfactory tracts. The cerebral hemispheres have a pyriform shape and contact the cerebellum dorsally across more than 64% of the total width of the brain. However, in comparison with modern birds, the cerebral hemispheres are less expanded and do not surround or envelope the olfactory tracts: the pineal gland emerges wedged between the lobes. The cerebellum is broad, being almost 50% of the total width of the brain. The optic lobes are large, globular to slightly pear-shaped, and laterally placed (and therefore well separated from each other); each is almost as large as the cerebellum. This pattern is shared with birds and, convergently, with pterosaurs^{15,18,19}. In modern birds, the optic lobes are rotated posteriorly around their basal long-axes and therefore partly overlap the medulla oblongata²⁰. In *Archaeopteryx* and *Enaliornis*²¹, an Early

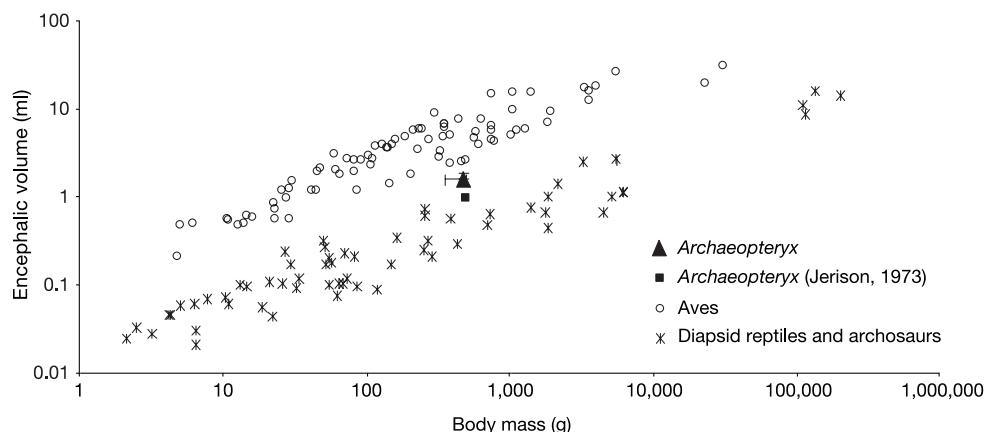


Figure 4 Encephalization index for birds, reptiles and BMNH 37001. Endoencephalic volumes of recent birds from ref. 17, brain masses for recent reptiles based on ref. 29; corrected to volumes using $\delta = 1.036 \text{ g cm}^{-3}$ (refs 3, 17). Mass for BMNH 37001 based

on mass = 468 g¹⁰, horizontal error bar indicates masses from 350 g to 500 g and the estimation of Jerison¹⁴. See also Table 1 in Supplementary Information.

Cretaceous bird, the optic lobes are similar in shape but are positioned more anteriorly. The cerebral, cerebellar and optic fossae are smooth and converge in the anterior part of the parietal bone, and the boundaries are noticeable externally. The pineal fossa is not visible externally. The cerebellar fossa has no traces of foliar structures, probably owing to the presence of the occipital dural venous sinus. The semicircular sulcus (the depression beside the crista arcuata) surrounds the floccular lobe placed just anterior to the opisthotic. The relatively small hemispheres, lack of posterior rotation of the optic lobes and potentially relatively smaller cerebellum suggest that *Archaeopteryx* and *Enaliornis*²¹ were more primitive than any modern bird.

The reconstructed 3D model of the brain encloses a volume of 1.4 ml. This must be considered as a minimum value because the base of the braincase is missing (around 88% of the total volume estimation). Previous estimates of relative brain size in *Archaeopteryx* varied by a factor of almost two, and it has remained contentious whether brain size increase was tied to the evolution of flight, arboreality, or other environmental influences¹⁴. We estimate the total volume to be around 1.6 ml, which is very near to Hopson's estimate of 1.76 ml (ref. 15). A plot of endocranial volume against body mass of some birds and modern reptiles (Fig. 4) places *Archaeopteryx* appreciably closer to birds than

Jerison's estimate does¹⁴. Birds with the same body mass as *Archaeopteryx* have from one-third (for example, galliforms and columbiforms) to five times (for example, psittaciforms and passeriforms) bigger brains^{14,15}. However, the brain of *Archaeopteryx* is about three times the volume of those of non-avian reptiles of equivalent size.

The right inner ear (Fig. 5) housed inside the right prootic and opisthotic elements is the better preserved, despite post-mortem rotational dislocation of both bones. The semicircular canals are not complete but their course and proportions can be established. The position of the anterior and lateral canals follows the archosaur pattern²². However, the anterior canal is considerably elongated and reflexed, and the course of the posterior canal extends ventrally below the plane of the lateral canal as in birds^{22,23}. Moreover, the plane of the posterior canal is almost transverse at the common crus, which is therefore in a more posterior position than in non-avian archosaurs and non-archosaur reptiles^{22,23}. The curved structure adjacent to the cochlea may represent the merged area of the vagal canal and metotic fissure.

We have plotted data for semicircular canal proportions and cochlea length (Fig. 6a, b) for selected birds, archosaurs and non-archosaur reptiles that place *Archaeopteryx* close to or within the range for modern birds (see Supplementary data), suggesting that

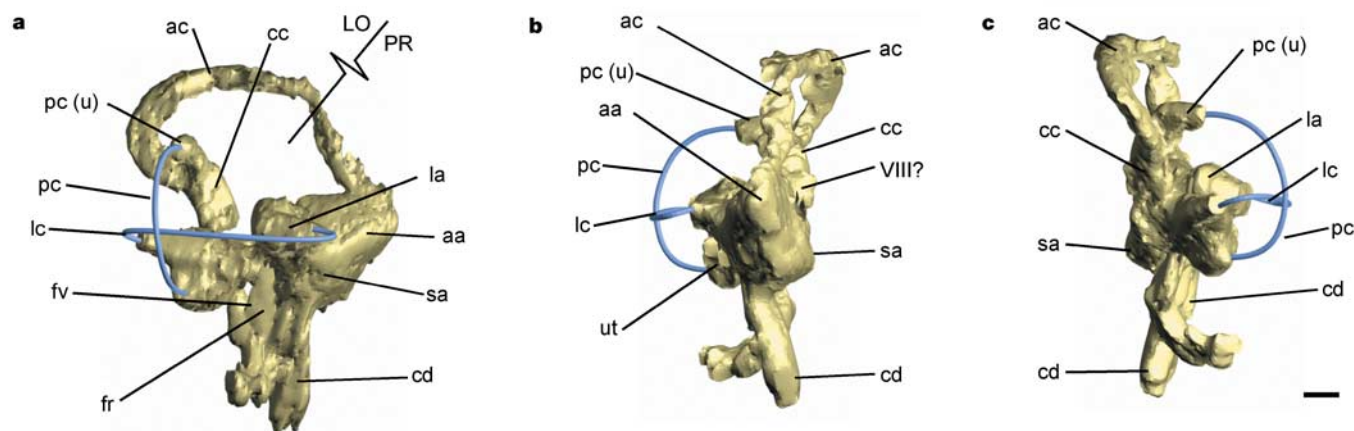


Figure 5 Right inner ear of BMNH 37001. 3D model based on X-ray CT. Scale bar, 2 mm. **a**, Lateral (external) view. **b**, Anterior view. **c**, Posterior view. Abbreviations: aa, anterior ampulla; ac, anterior canal; cc, common crus; cd, cochlear duct; fr, fenestra pseudorotunda; fv, fenestra vestibuli (fenestra ovalis); la, lateral ampulla; lc, tentative

course of lateral canal; LO, left opisthotic; pc, tentative course of posterior canal; pc (u), posterior canal (upper part); PR, prootic; sa, sacculus; ut, utricle; VIII?, acoustic nerve or endolymphatic duct. Original imagery available at < <http://www.DigiMorph.org> > .

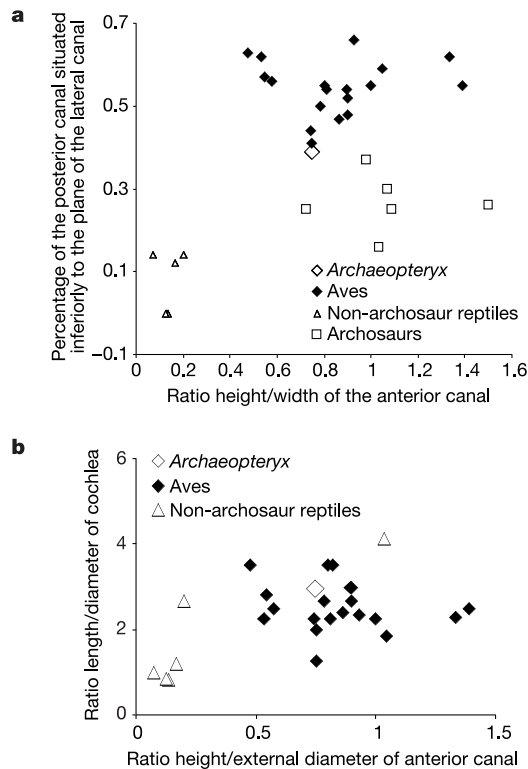


Figure 6 Comparative proportions of the inner ear of BMNH 37001, selected recent birds, archosaurs and non-archosaur reptiles. **a**, Anterior and posterior semi-circular canals. **b**, Length of the cochlea. All data on modern taxa were taken from ref. 23 (see also Supplementary Information) except *Alligator*²². For sources of data on fossil taxa see Table 2 in Supplementary Information.

the spatial sensory perception and auditory abilities of *Archaeopteryx* were similarly developed. The cochlea duct elongates in modern birds to enhance hearing ability²⁴ and offers an indication of the role of sonic stimuli in behavioural responses. This is much greater than in many other vertebrates and is associated with territorial, mating and feeding behaviour²⁵.

Few coelurosaurian dinosaur taxa preserve a brain endocast and most show the non-flexed plesiomorphic brain pattern with the optic lobes in dorsal position^{18,26}, except for *Troodon*¹⁸ in which they are laterally placed; however, the cerebellum is comparatively much smaller than in *Archaeopteryx*. Maniraptorans do, however, show a trend towards brain enlargement and laterally separated optic lobes²⁷, with *Archaeopteryx* exhibiting a 'flight adapted' brain, a stage further towards the modern bird pattern. The remodelling of the brain towards the avian condition must have begun well before the appearance of *Archaeopteryx* 147 million years ago in the latest Jurassic. The convergent increase in visual and vestibular regions in pterosaurs¹⁹ is further evidence that both an aerodynamic wing and a powerful central nervous system are integral to powered flight. □

Methods

The braincase of BMNH 37001 was scanned at the University of Texas at Austin's High-Resolution X-ray CT Facility. The specimen was scanned twice, at low and high X-ray energies (120 kV and 180 kV, respectively); the former accentuates compositional differences, whereas the latter is more noise-free and less prone to beam-hardening artefacts and interference from high-attenuation phases. The resulting serial sections were saved as two independent series of 1313 16-bit TIFF files. After a preliminary exploration, every second section of the low-energy data set was used to create 3D computer models of the objects by reconstructing the surfaces that connected corresponding outlines on adjacent sections²⁸. The resulting matrix data size was 1,024 × 1,024 × 650 voxels and the voxel size was 20 × 20 × 46 μm. The data set was segmented using Mimics v7.3 software (Materialise). Because of lateral changes and similarity in the X-ray attenuation values

within the bone and limestone matrix, and the presence of highly attenuating manganese dioxide crystals, it was necessary to use local thresholds and work in the three orthogonal planes checking continuously the temporal 3D models of each bone as they were built up. Ogle (M. J. Gourlay, Northwestern Research Associates Inc.) and Simian v2 (University of Utah/Los Alamos National Laboratory) were employed for supplementary volume-rendering analysis.

The parameters common to both scan series were: X-ray spot size 0.030 mm; 1,024 detector channels; slice thickness and inter-slice spacing 0.0230 mm; field of reconstruction 21.0 mm; source to object distance 67 mm; source to detector distance 596 mm; 2,000 views; 200 ms acquisition time per view; acquired with 21 slices per rotation. Ring artefacts were corrected by R.A.K. using an original algorithm that processed the raw CT sinogram data to remove radially consistent channel-to-channel divergences, and resulted in no introduction of spurious features.

Received 1 March; accepted 28 May 2004; doi:10.1038/nature02706.

- Ostrom, J. H. *Archaeopteryx* and the origin of birds. *Biol. J. Linn. Soc.* **8**, 91–182 (1976).
- Ji, Q., Norell, M. A., Gao, K. Q., Ji, S. A. & Ren, D. The distribution of integumentary structures in a feathered dinosaur. *Nature* **410**, 1084–1088 (2001).
- Padian, K. & Chiappe, L. M. The origin and early evolution of birds. *Biol. Rev.* **73**, 1–42 (1998).
- Shimizu, T. in *Brain Evolution and Cognition* (eds Roth, G. & Wullman, M. F.) 135–184 (Wiley, New York, 2001).
- Ketcham, R. A. & Carlson, W. D. Acquisition, optimization and interpretation of X-ray computed tomographic imagery applications to the geosciences. *Comp. Geosci.* **27**, 381–400 (2001).
- Whybrow, P. J. Preparation of the cranium of the holotype of *Archaeopteryx lithographica* from the collections of the British Museum (Natural History). *Neues Jb. Geol. Palaeont. Mh.* **3**, 184–192 (1982).
- Whetstone, K. N. Braincase of Mesozoic birds: 1. New preparation of the 'London' *Archaeopteryx*. *J. Vert. Paleol.* **2**, 439–452 (1983).
- Walker, A. in *The Beginnings of Birds. Proceedings of the International Archaeopteryx Conference, Eichstatt 1984* (eds Hecht, M. K., Ostrom, J. H., Wellnhofer, P. & Viohl, G.) 123–134 (Freunde des Jura-Museums Eichstatt, Eichstatt, 1985).
- Elzanowski, A. & Wellnhofer, P. Cranial morphology of *Archaeopteryx*: evidence from the seventh skeleton. *J. Vert. Paleol.* **16**, 81–94 (1996).
- Elzanowski, A. in *Mesozoic Birds. Above the Heads of Dinosaurs* (eds Chiappe, L. M. & Witmer, L. M.) 129–159 (Univ. of California Press, Berkeley, 2002).
- Witmer, L. M. The craniofacial air sac system of Mesozoic birds (Aves). *Zool. J. Linn. Soc.* **100**, 327–378 (1990).
- Witmer, L. M. in *Encyclopedia of Dinosaurs* (eds Currie, P. J. & Padian, K.) 151–159 (Academic, San Diego, 1997).
- Jerison, H. J. Brain evolution and *Archaeopteryx*. *Nature* **219**, 1381–1382 (1968).
- Jerison, H. J. *Evolution of the Brain and Intelligence* (Academic, New York, 1973).
- Hopson, J. Relative brain size and behaviour in archosaurian reptiles. *Annu. Rev. Ecol. Syst.* **8**, 429–448 (1977).
- Stark, D. in *Biology of the Reptilia* Vol. 9 Neurology A (eds Gans, C., Northcutt, R. G. & Ulinsky, P.) 1–38 (Academic, London, 1979).
- Iwaniuk, A. N. & Nelson, J. E. Can endocranial volume be used as an estimate of brain size in birds? *Can. J. Zool.* **80**, 16–23 (2002).
- Hopson, J. A. in *Biology of the Reptilia* Vol. 9 Neurology A (eds Gans, C., Northcutt, R. G. & Ulinsky, P.) 39–147 (Academic, London, 1979).
- Witmer, L. M., Chatterjee, S., Franzosa, J. & Rowe, T. Neuroanatomy of flying reptiles and implications for flight, posture and behaviour. *Nature* **425**, 950–953 (2003).
- Dubbeldam, J. L. in *The Central Nervous System of Vertebrates* (eds Nieuwenhuys, R., ten Donckelaar, H. J. & Voogd, J.) Vol. 3 1525–1636 (Springer, Berlin, 1998).
- Elzanowski, A. & Galton, P. M. Braincase of *Enaliornis*, an Early Cretaceous bird from England. *J. Vert. Paleol.* **11**, 90–107 (1991).
- Wever, E. G. *The Reptile Ear. Its Structure and Function* (Princeton Univ. Press, Princeton, 1978).
- Gray, A. A. *The Labyrinth of Animals Including Mammals, Birds, Reptiles and Amphibians* Vol. II (J & A Churchill, London, 1908).
- Gleich, O. & Manley, G. A. in *Comparative Hearing: Birds and Reptiles* (eds Dooling, R. J., Fay, R. R. & Popper, A. N.) 70–138 (Springer, New York, 2000).
- Pearson, R. *The Avian Brain* (Academic, London, 1972).
- Brochu, C. A. Osteology of *Tyrannosaurus rex*: insights from a nearly complete skeleton and high-resolution computed tomographic analysis of the skull. *Mem. Soc. Vert. Paleol.* **7**, 1–138 (2003).
- Franzosa, J. *Evolution of the Brain in Theropoda (Dinosauria)* PhD dissertation, Univ. Texas (2004).
- Dominguez, P., Jacobson, A. G. & Jefferies, R. P. S. Paired gill slits in a fossil with a calcite skeleton. *Nature* **417**, 841–844 (2002).
- Platel, R. in *Biology of the Reptilia* (eds Gans, C., Northcutt, R. G. & Ulinsky, P.) Vol. 9 Neurology A 147–171 (Academic, London, 1979).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank UTCT and NSF for funding CT scanning, and M. Colbert and J. Maisano (UTCT) for technical assistance. L. M. Witmer provided constructive comment and criticism. NERC funded the project, and the NHM Palaeontology Department Research Fund provided supplementary funds for P.D.A. We dedicate this paper to the memory of P. J. Whybrow, formerly Head of the Palaeontology Laboratory at the Natural History Museum London, who died suddenly on 13 February 2004. His skilled preparation work, undertaken long before the advent of CT technology, made this project possible.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.C.M. (A.Milner@nhm.ac.uk).

Processing of wild cereal grains in the Upper Palaeolithic revealed by starch grain analysis

Dolores R. Piperno^{1,2*}, Ehud Weiss^{3*}, Irene Holst¹ & Dani Nadel⁴

¹Smithsonian Tropical Research Institute, Box 2072, Balboa, Republic of Panama

²Archaeobiology Program, Department of Anthropology, National Museum of Natural History, Smithsonian Institution, Washington DC 20560, USA

³Department of Anthropology, Peabody Museum, Harvard University, 11 Divinity Avenue, Cambridge, Massachusetts 02130, USA

⁴Institute of Archaeology, University of Haifa, Mount Carmel, Haifa 31905, Israel

* These authors contributed equally to this work.

Barley (*Hordeum vulgare* L.) and wheat (*Triticum monococcum* L. and *Triticum turgidum* L.) were among the principal 'founder crops' of southwest Asian agriculture¹. Two issues that were central to the cultural transition from foraging to food production are poorly understood. They are the dates at which human groups began to routinely exploit wild varieties of wheat and barley, and when foragers first utilized technologies to pound and grind the hard, fibrous seeds of these and other plants to turn them into easily digestible foodstuffs. Here we report the earliest direct evidence for human processing of grass seeds, including barley and possibly wheat, in the form of starch grains recovered from a ground stone artefact from the Upper Palaeolithic site of Ohalo II in Israel. Associated evidence for an oven-like hearth was also found at this site, suggesting that dough made from grain flour was baked. Our data indicate that routine processing of a selected group of wild cereals, combined with effective methods of cooking ground seeds, were practiced at least 12,000 years before their domestication in southwest Asia.

Ohalo II is located on the southwestern shore of the Sea of Galilee at an elevation of 212 m below sea level. In most years the site is submerged in 2–3 m of water. Ohalo II was exposed in 1989 and 1999, following a dramatic drop in the water level of the lake; today the site covers an area of about 2,000 m². Three seasons of excavations took place during each of the low-level events before it was submerged again. The work revealed the remains of a camp, including six brush huts built from branches and leaves, hearths and a human grave^{2,3}. All the features were well-preserved, having been found *in situ* in the waterlogged sediments. A series of 40 ¹⁴C determinations on charred plant remains dates the occupation to around 19,500 ¹⁴C yr BP (about 23,500 to 22,500 calendar years ago) placing the site within the Upper Palaeolithic period (locally also termed the Early Epipalaeolithic)⁴.

Exceptional preservation of organic material in the fine lake sediments enabled the retrieval of rich faunal (mammals such as gazelles, birds, rodents, fish and molluscs) and plant assemblages (more than 30 economic species), including numerous charred grains of wild barley (*Hordeum spontaneum* Koch) and emmer wheat (*Triticum dicoccoides* (Ascherson and Graebner) Aaronson) and other grass species (for example, *Bromus pseudobrachystachys* Hornug and *Bromus tigridis* Boiss and Noe)^{5,6}. These remains represent the earliest evidence for considerable use of grasses in the human diet.

Hut 1 at Ohalo II is the oldest and best preserved among the huts recovered to date, and three successive floors and wall bases were identified and described. A large, flat basalt stone was set on a patch of sand on the second floor supported by small pebbles, much like an anvil or a work surface (Fig. 1). Discrete concentrations of food and medicinal plant seeds were distributed on the floor around the stone. To ascertain whether the Hut I stone was used to grind the plants, and specifically to establish which taxa were processed, we

carried out starch grain studies on this artefact. To recover starch grains from the stone, we used a protocol shown to be successful with large stone artefacts in other areas of the world^{7,8} (see Methods). We developed a key for starch grain identifications on the basis of a large modern reference collection that included many grasses, other economic plants and weeds native to the study region. Plant taxa that were previously identified at Ohalo II^{5,6} were given close attention (see Methods). We found that the genera *Hordeum*, *Triticum* and *Aegilops* (hereafter, the AHT taxa group) are separable from other grasses native to the study region on the basis of morphological and size criteria. Furthermore, specific grain types as well as grain population signatures can be used to identify the presence and probable absence or low frequency of each of these genera in archaeological starch assemblages (Fig. 2a–g and Table 1a) (see also Supplementary Figs 1–7 for detailed descriptions and illustrations of modern starch grains).

A total of 150 starch grains were recovered from the grinding stone. A hundred and twenty-seven of these were identified as grass seed starches. Seventy-eight out of those 127 grains are diagnostic of the AHT taxa group on the basis of their lenticular shapes and sizes (Table 1b and Fig. 3a). A genus-specific identification of starch grains from *Hordeum* is possible because they commonly (about 50% of all the lenticular forms present in the starch population) have well-defined lamellae that are differentiable from lamellae in *Aegilops*; *Triticum* grains have no such demonstrable lamellae (Fig. 2a–d and Supplementary Figs 1–3). On the basis of this we identified 23 grains from *Hordeum* on the stone (Fig. 3b, c). Although the size characteristics of the *Hordeum* grains with lamellae are most similar to modern *Hordeum spontaneum* (Table 1a, b), distinguishing between *H. spontaneum*, *Hordeum glaucum* and *Hordeum bulbosum* is not possible at this time.

Hordeum marinum, the carbonized seeds of which were present in appreciable quantities around the stone (a total of 505 seeds) can, however, be ruled out from representation in the starch grain residues from the stone because starches from this species have much smaller sizes and a characteristic morphology (Supplementary Fig. 4). We found no grains with the crater-like depressions that commonly occur in populations of starches from *Triticum* and may be diagnostic of that genus (Fig. 2a). Similarly, whereas many grains from modern *Aegilops* spp. have central swellings and/or prominent fissures together with distinctive lamellae (Fig. 2c, d), no grains of that type were observed. In view of the large sample size of grains recovered from the stone, and the reasonable expectation that if *Triticum* or *Aegilops* had been processed with any frequency then grains characteristic of those genera just described should have been



Figure 1 The stone implement analysed from Ohalo II.

recovered, the majority of the grains on the stone are probably from *Hordeum*.

The 49 other Poaceae grains retrieved from the stone are compound grains without characteristic features, and can belong to any number of genera in the family, including the AHT group (Fig. 3d). Importantly, however, we could find no evidence that grass seeds from genera outside the AHT group, or other non-grass species that were commonly to moderately present as carbonized remains from the hut floor, had been processed. Starch grains from these species possess markedly different shape and surface features when compared with those recovered from the grinding stone (see Table 1a, Methods and Supplementary Figs 5–7). No starches characteristic of roots and tubers were present in the tool residues. Hence, the stone seems to have been a speciality, cereal-processing implement.

A special alignment of burned stones covered by ash at Ohalo II suggests the presence of a hearth-like structure used as a simple oven; if this assumption is correct it would represent the earliest such food-roasting strategies documented. A few metres southwest of Hut 1, 13 natural and worked stones were arranged together to form a paved circle (Supplementary Fig. 8). Above and between the stones was a thin layer of dark material rich in charcoal and ash but

with few carbonized fruits and grains, mostly from *H. spontaneum*. Other hearths at the site were not lined by or paved with any stones. They were visible instead as simple, round or oval ash deposits, usually 30–50 cm across and several centimetres thick. We believe that the stone-paved feature was used for baking food. The persistence and use of similar kinds of simple ovens by recent and modern nomads and hunter-gatherers provides an ethnographic analogy for the Ohalo II example. Groups in southwest Asia, Sahara and Australia prepare rough dough from grass and other ground seeds by mixing flour and water in a container or even on a piece of cloth, and then placing the dough on top of a layer of heated stones from which the embers have been removed. After the dough is spread on the stones, the embers are put back to cover the dough^{9–11}.

The beginning of baking was probably an important step forward in human nutrition. For example, the glycemic index reflects how rapidly glucose from carbohydrates enters the blood after ingestion, and it is influenced considerably by methods of food preparation that involve particle size reduction and mixing of starches with water, and by the duration and manner of cooking^{12,13}. Ingesting ground and baked barley and wheat products would have had the effect of increasing the glycemic index (and thus the amount of available dietary energy) by between 56% and 72%, compared with ingesting uncooked or even cooked whole-grained forms of these plants (Supplementary Table 1). Although no baked product was preserved at Ohalo II, the starch grain data together with the fact that few carbonized grains were found in the oven feature suggest that what was roasted included dough prepared from wild cereals.

If they are going to serve as adequate dietary staples, grasses, other seeds and underground plant organs usually must be pounded and ground to separate and remove indigestible fibrous substances and toxins, reduce the size of food particles, and concentrate nutritious, high-energy starchy reserves¹⁴. The point at which technological improvements such as these began to be employed in plant food preparation, and which taxa and their specific structures (for example, roots, tubers, seeds and nuts) were first processed in this way are old and important questions. In southwest Asia, Europe and northern Africa, ground stone implements of various types, identifiable as possible plant grinders or pounders (for example, hammerstones grinding slabs, mortars and pestles), make their first appearance during the Upper Palaeolithic between around 45,000 and 18,000 BP, and they were manufactured by fully-modern humans^{15,16}. However, because once-associated macroscopic plant remains have rarely survived along with these artefacts and because

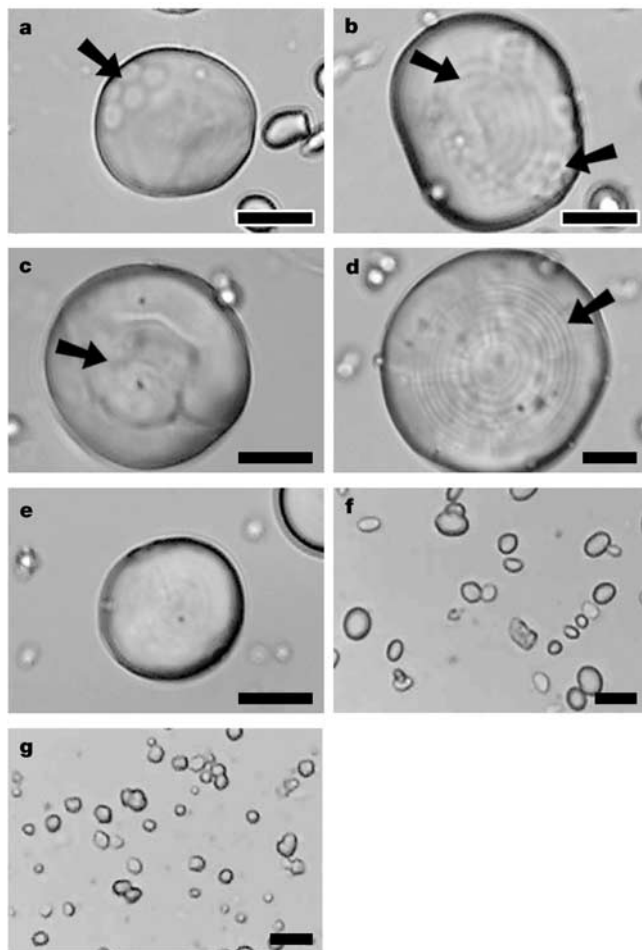


Figure 2 Characteristic starch grains from seeds of various modern cereals and other grasses. **a**, *T. dicoccoides*. The arrow on the left points to distinctive crater-like depressions. **b**, *H. spontaneum*. The arrow on the left points to lamellae; arrow on the right points to *Hordeum*-type surface depressions. **c**, *Aegilops geniculata*. The arrow points to the central protuberance with the pleat. **d**, *Aegilops peregrina*. The arrow points to the characteristic lamellae present in this species. **e**, *H. glaucum*. **f**, *B. pseudobrachystachys*. **g**, *Piptatherum holciforme*. Scale bars, 10 μ m.

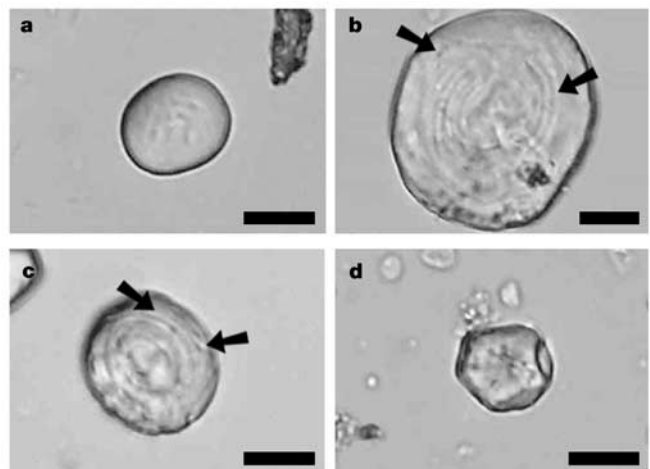


Figure 3 Starch grains recovered from the stone implement. **a**, Lenticular grain from the AHT taxa group. **b**, **c**, Lenticular grains with lamellae (right arrow in **b** and left arrow in **c**) from *Hordeum*. They also have surface depressions (other arrows) characteristic of the genus. **d**, Compound starch grain. Scale bars, 10 μ m.

Table 1 Starch grain size attributes of modern and archaeological grasses

	Mean ($\pm$ s.d.) of length (μ m)	Range	n
(a) Analysed modern grasses from the study region			
<i>Aegilops geniculata</i> Roth	21 $\pm$ 6.4	10–36	50
<i>Aegilops peregrina</i> (Hackel) Maire and Weiller	25 $\pm$ 8.0	12–52	50
<i>Hordeum spontaneum</i> Koch*	20 $\pm$ 4.7	10–26	50
<i>Hordeum bulbosum</i> L.	17 $\pm$ 3.7	10–24	50
<i>Hordeum glaucum</i> Steudel*	18 $\pm$ 3.9	8–24	50
<i>Hordeum marinum</i> Hudson*	10 $\pm$ 1.8	6–14	50
<i>Triticum dicoccoides</i> (Ascherson and Graebner) Aaronsohn	17 $\pm$ 6.1	8–30	50
<i>Hordeum bulbosum</i> with lamellae only	21 $\pm$ 1.6	18–24	50
<i>Hordeum glaucum</i> with lamellae only	22 $\pm$ 1.4	18–26	50
<i>Hordeum spontaneum</i> with lamellae only	28 $\pm$ 2.9	22–40	50
<i>Alopecurus utriculatus</i> Banks and Solander	5 $\pm$ 1.5	2–8	50
<i>Alopecurus arundinaceus</i> L.	4 $\pm$ 0.9	2–8	50
<i>Avena barbata</i> Pott and Link	12 $\pm$ 2.9	6–18	50
<i>Brachypodium distachyon</i> (L.) Palisot de-Beauvois	9 $\pm$ 2.2	4–16	50
<i>Bromus pseudobrachystachys</i> Hornug*	5 $\pm$ 1.4	4–8	50
<i>Gastridium ventricosum</i> (Gouan) Schinz and Thellung	4 $\pm$ 1.0	2–6	50
<i>Lolium multiflorum</i> Lamarck	<6.0	–	50
<i>Lolium rigidum</i> Gaudin	<6.0	–	50
<i>Phalaris minor</i> Retzius	<2.0	–	50
<i>Phalaris paradoxa</i> L.	<4.0	–	50
<i>Puccinellia gigantea</i> (Grossheim) Grossheim	<4.0	–	50
<i>Puccinellia distans</i> (L.) Parlatores	<4.0	–	50
<i>Piptatherum holciforme</i> * (Bieberstein) Roemer and Schultes	3 $\pm$ 1.0	2–4	50
<i>Vulpia persica</i> Boissier	<2.0	–	50
(b) Archaeological Poaceae starch grains†			
All Poaceae	20 $\pm$ 6.6	8–40	98
Lenticular grains (AHT taxa)	23 $\pm$ 6.3	14–40	54
<i>Hordeum</i> grains with lamellae only	28 $\pm$ 4.7	20–38	23

* Carbonized remains of these species were common around the grinding stone and on the floor of Hut 1.

† Seven of the starch grains were recovered from the 10 needle probes of small cavities in the stone. The remainder of the sample of starches, 143 grains, resulted from Step II (high energy shaking) (see Methods). Step II yielded very little sediment in association with the starch grains. These results indicate that most of the grains recovered had been lodged inside cracks and crevices on the stone, where they were protected from degradation and decay.

some of the implements contain the remains of ochre, distinguishing between plant food preparation and the grinding of mineral pigments (or even animal bone) as their primary function, together with determining which plants may have been processed, have been difficult.

Our findings provide direct empirical data that grinding stones from the Upper Palaeolithic were used to process plant foods, and they constitute the earliest evidence for this activity. Our data indicate that a selected cadre of wild cereals, including barley and possibly also wheat and/or goat grass—the first two of which were subsequently domesticated between around 10,000 and 9,000 BP—received preferential emphasis for processing, and may thus have contributed significantly to the human diet as far back as around 20,000 BP. Wild cereals have the largest and densest grains of any grasses native to southwest Asia, an attribute that may have caused them not only to be attractive to early seed collectors¹⁷ but also to be favoured for intensive methods of processing once gathered. Smaller-seeded grasses may simply have been parched before consumption⁶. Furthermore, grinding the largest-grained grasses may have ameliorated the high ‘handling costs’ and decreases of foraging efficiency associated with the shift to broader spectrum subsistence and incorporation of intensive grass seed exploitation strategies evidenced at Ohalo II^{18–21}. This would have further encouraged close relationships between foragers and the closest wild relatives of domesticated cereals. Nutritionally effective ways of preparing foods from wild cereals, evidenced at Ohalo II, likely facilitated other forms of subsistence intensification (such as seed storage²²) that were preludes to domestication and became established in the region during the following 10,000 yr.

Our research opens up a new avenue of palaeoethnobotanical investigation in southwest Asia and other areas of the world where ground stone tool assemblages were used before the Neolithic period. Future starch grain studies have the potential to elucidate when and where grasses, other seeds and nuts, and underground plant organs first became important components of human diets and technological efforts. The functions of ground (and chipped) stone tools recovered from early archaeological sites can also be

clarified by such research. In sites of any age from southwest Asia where other types of plant remains are not well-represented, starch grain data may stand alone as definitive evidence of cereal and other plant usage. □

Methods

Extraction of starch grains

The grinding stone was examined under a stereoscopic microscope at a magnification of $\times 100$. The point of a fine needle was inserted into cracks and crevices to loosen and remove any residue. Two probes, each at five different locations on the stone, were examined in this manner (step I). The residue was mounted in water on a slide and examined with polarized and unpolarized light at a magnification of $\times 400$. The stone was then shaken in an ultrasonic bath for 5 min to completely dislodge adhering sediment and starch. Starch was then isolated by adding a heavy liquid solution of caesium chloride (CsCl) with a density of 1.8 g cm^{-3} (step II). Starch grains were mounted on slides in water and examined microscopically as in after Step I. Sediments that originally surrounded the stone *in situ* were not available for analysis. All archaeological starch grains possessed the starch-specific extinction cross when examined under cross-polarized light.

Starch grain identification

We used a modern reference collection of over 400 different species from 45 different families of plants that is housed at the Smithsonian Tropical Research Institute. Included were 21 species of the Poaceae and 13 other species from seven different families whose carbonized remains were commonly retrieved and identified at Ohalo II, including around the grinding stone that was analysed (Table 1; *Galium tricornutum* Dandy, *Malva parviflora* L. and *Atriplex* spp. were the most numerous non-Poaceae taxa identified near the stone). The numerous studies available on starch grain morphology were also consulted^{7,8,23–28}. Our starch keys and classifications emphasize attributes shown to be useful in identifying individual families, genera and species of plants, including: (1) overall grain type (simple or compound) and shape (bell-shaped, circular, lenticular or oval), (2) contour and surface features, (3) position and form of the hilum (the botanical centre of the grain) and fissure (internal cracks emanating from the hila of some starch grains, formed when the grain begins to grow outward from the hilum and quite literally cracks), (4) number and characteristics of pressure facets present on compound grains and (5) presence or absence of demonstrable lamellae (rings formed during starch grain growth).

A considerable amount of research indicates that Poaceae seed starches are readily distinguishable from those of other families^{7,8,23–28}. We found that size alone effectively separates many wild grass species and genera of the study region from the AHT group, as mean sizes and ranges in the latter group far exceed those of other grasses studied (Table 1a). Importantly, morphological criteria also distinguish these grasses from the AHT group. AHT starches are usually simple (versus compound) grains, and circular, oval or kidney-shaped in outline, and when rotated into side view they assume a lenticular form (Fig. 2a–d and Supplementary Figs 1–4). In contrast, starch grains of other genera studied, including *Bromus* and *Piptatherum*, are either compound, flat and/or angular to

irregular in shape, and lack the features characteristic of the AHT starches (Fig. 2f, g).

It is also clear that using specific grain types together with attribute combinations in a multiple grain analysis is an effective and perhaps the most conservative means to distinguish individual genera and species. Such an approach employs all the morphological characteristics that account for the population of starches in a single species as well as their quantitative frequency tendencies, and it takes into account intra- and interspecific variation in grain attributes. For example, starch grain populations from *T. dicoccoides* contain high proportions (about 40%) of grains that have distinctive, large crater-like impressions on the surface; these grains are also without lamellae (Fig. 2a). These types were not observed in either *Hordeum* or *Aegilops*, which, in turn contributed high frequencies of other types of characteristic starch grains (Fig. 2b–e and Supplementary Figs 1–4). In archaeological starch grain assemblages of sufficient sample sizes it would be possible to identify the presence and/or probable absence or low frequency of individual genera and species using these kinds of signatures.

Other families and species represented through their carbonized remains at Ohalo II were: Asteraceae (*Centaurea pallescens* Delile, *Silybum marianum* (L.) Gaertner), Chenopodiaceae (*Atriplex halimus* L., *Suaeda aegyptiaca* (Hasselquist) Zohary), Fabaceae (*Melilotus indicus* (L.) Allioni, *Pisum elatius* Marschall von Bieberstein), Malvaceae (*Malva aegyptia* L.), Potamogetonaceae (*Potamogeton pectinatus* L., *Potamogeton perfoliatus* L.), Ruppiaceae (*Ruppia maritima* L.) and Zygophyllaceae (*Nitraria schoberi* L.). Most of these species produced oils and not starches. The taxa with starches (*Pisum*, *Potamogeton* and *Ruppia*) have grains that can be distinguished from others in our and other established reference collections (Supplementary Figs 5–7).

Received 4 May; accepted 4 June 2004; doi:10.1038/nature02734.

- Zohary, D. & Hopf, M. *Domestication of Plants in the Old World*, 3rd edn (Oxford Science Publications, Oxford, 2000).
- Nadel, D. *Ohalo II—A 23,000-Year-Old Fisher-Hunter-Gatherers' Camp on the Shore of the Sea of Galilee* (Catalogue No. 20, Hecht Museum, Univ. Haifa, Haifa, 2002).
- Nadel, D. & Werker, E. The oldest ever brush hut plant remains from Ohalo II, Jordan Valley, Israel. *Antiquity* **73**, 755–764 (1999).
- Nadel, D., Carmi, I. & Segal, D. Radiocarbon dating of Ohalo II: Archaeological and methodological implications. *J. Archaeol. Sci.* **22**, 811–822 (1995).
- Kislev, M. E., Nadel, D. & Carmi, I. Epipaleolithic (19,000 BP) cereal and fruit diet at Ohalo II, Sea of Galilee, Israel. *Rev. Palaeobot. Palynol.* **73**, 161–166 (1992).
- Weiss, E., Kislev, M. E., Simchoni, O. & Nadel, D. Small-grained wild grasses as staple foods at the 23,000 year old site of Ohalo II, Israel. *Econ. Bot.* (in the press).
- Piperno, D. R., Ranere, A. J., Holst, I. & Hansell, P. Starch grains reveal early root crop horticulture in the Panamanian tropical forest. *Nature* **407**, 894–897 (2000).
- Pearsall, D. M., Chandler-Ezell, K. & Zeidler, J. A. Maize in ancient Ecuador: Results of residue analysis of stone tools from the Real Alto site. *J. Archaeol. Sci.* **31**, 423–442 (2004).
- Palmer, C. Milk and cereals: identifying food and food identity among Fellahin and Bedouin in Jordan. *Levant* **34**, 173–195 (2002).
- Hobbs, J. J. *Bedouin Life in the Egyptian Wilderness* (Univ. of Texas Press, Austin, 1992).
- Nicolaïsen, J. *Ecology and Culture of the Pastoral Tuareg* (Copenhagen National Museum, Copenhagen, 1963).
- Jenkins, D. J. A. *et al.* Glycemic index: overview of implications in health and disease. *Am. J. Clin. Nutr.* **76**, 266–273 (2002).
- Foster-Powell, K., Holt, S. H. & Brand-Miller, J. C. International table of glycemic index and glycemic load values. *Am. J. Clin. Nutr.* **76**, 5–56 (2002).
- Stahl, A. in *Foraging and Farming* (eds Harris, D. R. & Hillman, G. C.) 171–194 (Unwin Hyman, London, 1989).
- Wright, K. Ground-Stone tools and hunter-gatherer subsistence in southwest Asia: Implications for the transition to farming. *Am. Antiq.* **59**, 238–263 (1994).
- Kuhn, S. L. & Stiner, M. C. in *Hunter-Gatherers: An Interdisciplinary Perspective* (eds Painter-Brick, C., Layton, R. H. & Rowley-Conwy, P.) 99–142 (Cambridge Univ. Press, Cambridge, 2001).
- Bar-Yosef, O. & Kislev, M. E. in *Foraging and Farming: The Evolution of Plant Exploitation* (eds Harris, D. R. & Hillman, G. C.) 633–642 (Unwin Hyman, London, 1989).
- Flannery, K. V. in *The Domestication and Exploitation of Plants and Animals* (eds Ucko, P. J. & Dingleby, G. W.) 73–100 (Duckworth, London, 1969).
- Hawkes, K., O'Connell, J. F. & Rogers, L. The behavioral ecology of modern hunter-gatherers, and human evolution. *Trends Ecol. Evol.* **12**, 29–32 (1997).
- Kennett, D. & Winterhalder, B. in *Foraging Theory and the Transition to Agriculture* (eds Kennett, D. & Winterhalder, B.) (Smithsonian Institution Press, Washington DC, in the press).
- Weiss, E., Wetterstrom, W., Nadel, D. & Bar-Yosef, O. The broad spectrum revolution revisited: Evidence from plant remains. *Proc. Natl Acad. Sci. USA* **101**, 9551–9555 (2004).
- Bar-Yosef, O. From sedentary foragers to village hierarchies: The emergence of social institutions. *Proc. Br. Acad.* **110**, 1–38 (2001).
- Reichert, E. T. *The Differentiation and Specificity of Starches in Relation to Genera, Species, Etc.* (Carnegie Institution of Washington, Washington DC, 1913).
- Seidemann, J. *Starke Atlas* (Paul Parey, Berlin, 1966).
- Cortello, A. R. & Pochettino, M. L. Starch grain analysis as a microscopic identification feature in the identification of plant material. *Econ. Bot.* **48**, 171–181 (1994).
- Perry, L. Starch analyses reveal the multiple functions of quartz 'manioc' grater flakes from the Orinoco Basin, Venezuela. *Interciencia* **27**, 635–639 (2002).
- Piperno, D. R. & Holst, I. The presence of starch grains on prehistoric stone tools from the lowland Neotropics: Indications of early tuber use and agriculture in Panama. *J. Archaeol. Sci.* **25**, 765–776 (1998).
- Torrence, R., Wright, R. & Conway, R. Identification of starch granules using image analysis and multivariate techniques. *J. Archaeol. Sci.* **31**, 519–532 (2004).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements Supported by the Smithsonian Tropical Research Institute (STRI), a grant to the STRI from the Andrew W. Mellon Foundation, the American School of Prehistoric Research (Peabody Museum), Harvard University and the National Museum of Natural History. We acknowledge the help of the following people and institutions for supplying the seed reference collection used in this study: E. Wood and D. Pfister (Harvard University Herbarium), the United States National Plant Germplasm System (North Central Regional Plant Introduction Station, Western Regional Plant Introduction Station and Plant Genetic Resources Conservation Unit), and the Royal Botanic Gardens, Kew (Seed Conservation Department). The Ohalo II project was supported by the Irene-Levi Sala CARE Archaeological project Foundation, the Israel Academy of Science, the Jerusalem Center for Anthropological Studies, the L.S.B. Leakey Foundation, the M. Stekelis Museum of Prehistory in Haifa, the MAFCAF Foundation, the National Geographic Society and the Israel Antiquities Authority.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.R.P. (pipernod@tivoliv.si.edu) and E.W. (eweiss@fas.harvard.edu).

Effect of trace metal availability on coccolithophorid calcification

K. G. Schulz^{1,2}, I. Zondervan¹, L. J. A. Gerringa³, K. R. Timmermans³, M. J. W. Veldhuis³ & U. Riebesell^{1,2}

¹Alfred Wegener Institute for Polar and Marine Research, PO Box 120161, 27515 Bremerhaven, Germany

²Leibniz Institute for Marine Sciences (IFM-GEOMAR), Düsterbrook Weg 20, 24105 Kiel, Germany

³Royal Netherlands Institute for Sea Research, PO Box 59, 1790 AB Den Burg, Texel, The Netherlands

The deposition of atmospheric dust into the ocean has varied considerably over geological time^{1,2}. Because some of the trace metals contained in dust are essential plant nutrients which can limit phytoplankton growth in parts of the ocean, it has been suggested that variations in dust supply to the surface ocean might influence primary production^{3,4}. Whereas the role of trace metal availability in photosynthetic carbon fixation has received considerable attention, its effect on biogenic calcification is virtually unknown. The production of both particulate organic carbon and calcium carbonate (CaCO₃) drives the ocean's biological carbon pump. The ratio of particulate organic carbon to CaCO₃ export, the so-called rain ratio, is one of the factors determining CO₂ sequestration in the deep ocean. Here we investigate the influence of the essential trace metals iron and zinc on the prominent CaCO₃-producing microalga *Emiliania huxleyi*. We show that whereas at low iron concentrations growth and calcification are equally reduced, low zinc concentrations result in a de-coupling of the two processes. Despite the reduced growth rate of zinc-limited cells, CaCO₃ production rates per cell remain unaffected, thus leading to highly calcified cells. These results suggest that changes in dust deposition can affect biogenic calcification in oceanic regions characterized by trace metal limitation, with possible consequences for CO₂ partitioning between the atmosphere and the ocean.

The production of CaCO₃ in the surface ocean, its export to greater depths and its deposition in the sediment above the lysocline (the depth below which CaCO₃ dissolves) affect atmospheric CO₂ in two ways. On a timescale that is shorter than the ocean mixing time, CaCO₃ export reduces alkalinity in the surface ocean and lowers its storage capacity for atmospheric CO₂. On a timescale of thousands of years, a mechanism called carbonate compensation balances CaCO₃ burial with its supply of raw materials (riverine calcium and carbonate ions) by adjusting the depth of the lysocline⁵. This determines the deep ocean's carbonate ion concentration and

controls surface water carbonate chemistry and consequently atmospheric CO_2 . On both timescales a reduction in CaCO_3 export lowers CO_2 concentrations in the atmosphere.

CaCO_3 is predominantly produced in the open ocean, to a large extent by coccolithophores⁶. These open-ocean areas are subject to strong changes in atmospheric trace metal deposition. The rates for both iron and zinc deposition are probably up to ten times lower in today's ocean compared with the glacial ocean (see ref. 1 and references therein for iron, and Fig. 1 for zinc). Today iron concentrations limit phytoplankton productivity in large parts of the ocean known as high nitrate low chlorophyll (HNLC) areas⁷. Outside these regions iron availability is generally sufficient for most primary producers; however, productivity is often limited by the macronutrient nitrogen⁸. Although the significance of open-ocean zinc limitation awaits further assessment, prevailing zinc concentrations have been found to limit phytoplankton growth⁹, particularly that of coccolithophores¹⁰.

To simulate open-ocean iron and zinc limitation and to examine their effect on biogenic calcification, the prominent CaCO_3 producer *E. huxleyi* was grown under well-defined trace metal conditions, using the trace metal chelator EDTA and varying additions of iron chloride (FeCl_3) or zinc chloride (ZnCl_2). Because *E. huxleyi* has the ability to substitute cobalt for zinc¹¹, cells were grown at low free cobalt concentrations ($[\text{Co}^{2+}]$) of about 0.2 pmol kg^{-1} , which are comparable to values found in open-ocean waters¹². Furthermore, because marine biogenic calcification is sensitive to CO_2 -related changes in seawater carbonate chemistry¹³, which has varied in parallel with atmospheric trace metal deposition¹, incubations were carried out at different CO_2 concentrations.

The effect of iron on calcification was tested in incubations covering a range of calculated $[\text{Fe(III)}']$ (total dissolved inorganic iron(III) species) between $0.06\text{--}360 \text{ pmol kg}^{-1}$, at CO_2 concentrations of 20 and $7 \text{ } \mu\text{mol kg}^{-1}$ (pH ~ 7.9 and ~ 8.35 , respectively). The instantaneous growth rate decreased with decreasing $[\text{Fe(III)}']$ and was half-saturated at $\sim 0.7 \text{ pmol kg}^{-1}$ (Fig. 2a). Over the observed range neither the CaCO_3 content per cell (Fig. 2b) nor the cellular CaCO_3 to nitrogen ratio (Fig. 2d) changed significantly; however, the CaCO_3 production rate per cell decreased about sixfold with decreasing $[\text{Fe(III)}']$ (Fig. 2c), proportionally with the growth rate.

In a second experiment the effect of zinc on calcification was studied. Four different $[\text{CO}_2]$ of 30, 20, 13 and $7 \text{ } \mu\text{mol kg}^{-1}$ (pH

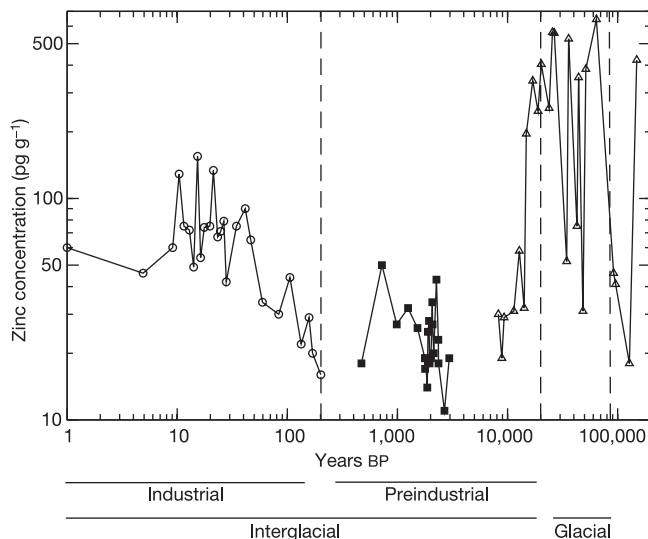


Figure 1 Zinc concentration in central Greenland snow and ice over the past 150,000 years. Data compiled from previous research: ref. 28, circles; ref. 29, squares; and ref. 30, triangles.

~ 7.75 , ~ 7.9 , ~ 8.1 and ~ 8.35 , respectively) were applied and the calculated free zinc concentration ($[\text{Zn}^{2+}]$) ranged from 0.3 to 6 pmol kg^{-1} . The instantaneous growth rate decreased with decreasing $[\text{Zn}^{2+}]$ and was half-saturated at $\sim 0.5 \text{ pmol kg}^{-1}$ (Fig. 3a), in close agreement with the previously determined value of $\sim 0.6 \text{ pmol kg}^{-1}$ (ref. 11). At the same time, the amount of CaCO_3 per cell increased from 10 to almost 60 pg with decreasing $[\text{Zn}^{2+}]$ (Fig. 3b). This was confirmed by scanning electron microscopy, which showed that cells of *E. huxleyi* were covered with multiple layers of calcite platelets (coccoliths) under zinc-depleted conditions, compared with 1–2 layers of coccoliths at high $[\text{Zn}^{2+}]$ (Fig. 4a and b, respectively). The ratio of cellular CaCO_3 to nitrogen increased more than twofold, from about four at high $[\text{Zn}^{2+}]$ to approximately nine at low $[\text{Zn}^{2+}]$ (Fig. 3d). In contrast, the CaCO_3 production rate per cell was hardly affected (Fig. 3c).

In the iron experiment, higher CaCO_3 contents and production rates per cell were observed at low $[\text{CO}_2]$, consistent with previous findings¹⁴ (Fig. 2b, c). Owing to the massive zinc-related response of the cellular CaCO_3 content (Fig. 3b) and the comparatively high variability within and between CO_2 treatments, the relatively small effect of carbonate chemistry on calcification could not be detected in the zinc assay.

The observed differences in calcification between iron- and zinc-limited cells may reflect the different roles of the two micronutrients in cellular processes. Although zinc deficiency slows down cellular growth and nitrogen utilization rates, it does not significantly affect the rate of CaCO_3 production (Fig. 3c). The resulting de-coupling of growth and calcification causes CaCO_3 accumulation in slow-growing cells of *E. huxleyi* (Fig. 3b). This also leads to a steady increase in the cellular CaCO_3 to nitrogen ratio when changing from high to low $[\text{Zn}^{2+}]$ (Fig. 3d). On the contrary, iron depletion not only reduces growth and nitrogen utilization rates, but equally reduces the rate of cellular CaCO_3 production (Fig. 2c). Therefore, the cellular CaCO_3 to nitrogen ratio remains fairly constant with varying iron availability (Fig. 2d).

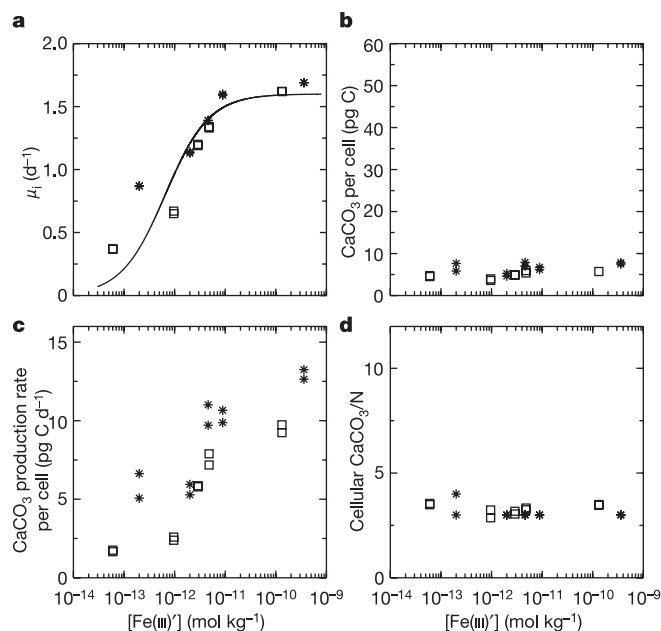


Figure 2 Response of *E. huxleyi* to varying free iron concentrations ($[\text{Fe(III)}']$). Responses shown are instantaneous growth rate (μ) (a), CaCO_3 content per cell (b), the specific CaCO_3 production rate per cell and day (c) and the ratio of cellular CaCO_3 to nitrogen (d). The solid line in a was obtained by fitting the data nonlinearly to a modified Monod curve, which accounts for a cell's minimum free metal requirement. The squares and asterisks denote CO_2 concentrations of 20 and $7 \text{ } \mu\text{mol kg}^{-1}$, respectively.

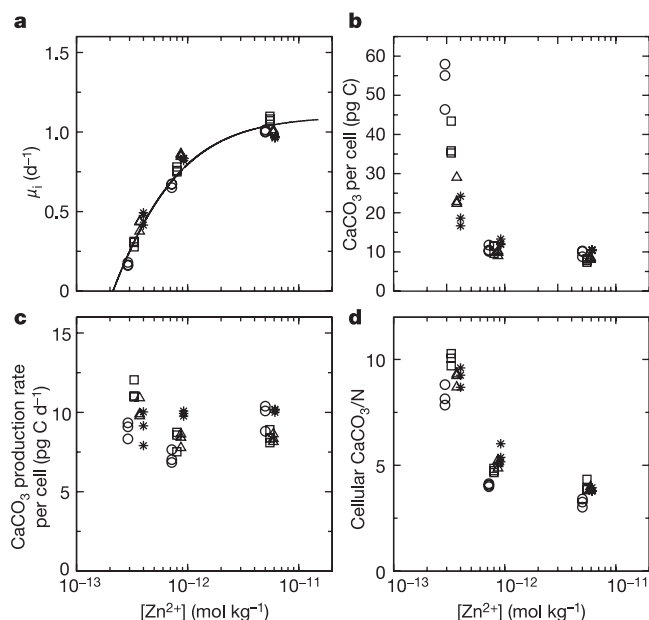


Figure 3 Response of *E. huxleyi* to varying free zinc concentrations ($[Zn^{2+}]$). **a–d**, As Fig. 2. The symbols denote four CO_2 concentrations as follows: circles, 30; squares, 20; triangles, 13; and asterisks, 7 $\mu mol\ kg^{-1}$. Note that the lower maximum growth rate in this experiment compared with the iron experiment (Fig. 2a) was due to the lack of the vitamin thiamine. A control experiment with thiamine added at selected $[Zn^{2+}]$ (data not shown) resulted in the same trends at higher μ_i .

The biogeochemical relevance of the observed responses to iron and zinc limitation is likely to differ between oceanic regions. Increased atmospheric deposition of iron and zinc in the ocean's HNLC areas during glacial times has probably not significantly altered global $CaCO_3$ production. For reasons not fully understood, pelagic calcification is only of minor importance in these regions. In the largest HNLC region, the Southern Ocean, $CaCO_3$ production seems to be negligible⁶. This is also true for the equatorial Pacific, which is dominated by siliceous primary production, as is clearly reflected in the underlying sediments¹⁵ above the lysocline. Only in the iron-deficient Pacific above 45° N does $CaCO_3$ production match that in the rest of the North Pacific⁶. As this area is comparatively small, however, its contribution to global calcification is bound to be low. Additionally, in HNLC areas diatoms have been shown to benefit most from iron or combined iron and zinc additions (see refs 9, 10 and references therein) by increasing nitrogen utilization and biomass production.

Outside the HNLC areas, export production is limited by the supply of macronutrients, primarily nitrate⁸. In many of these areas total dissolved zinc concentrations (see refs 16, 17 and references therein) are as low as those found in parts of the North Pacific¹⁸, where coccolithophorid growth has been shown to be limited by zinc¹⁰. Considering that in the present ocean up to 450 $\mu g\ m^{-2}$ dissolvable zinc is deposited atmospherically per year¹⁹ compared with 1,000 $\mu g\ m^{-2}$ zinc per year which is upwelled (taking a deep-ocean zinc concentration of 5.4 $nmol\ kg^{-1}$ and a mean deep-surface water exchange volume of 3 $m^3\ m^{-2}$ per year¹⁵), a tenfold increase (Fig. 1) in atmospheric deposition can be expected to significantly alter surface ocean zinc concentrations. Moreover, Northern Hemisphere atmospheric zinc deposition in the recent pre-industrial period was probably one-half to one-third of that today, as indicated by zinc measurements taken in Arctic ice and snow (Fig. 1). Thus, regions in which growth of coccolithophores is potentially zinc-limited today probably experienced significant alleviation during glacial times.

There are two ways in which changes in zinc supply to the surface

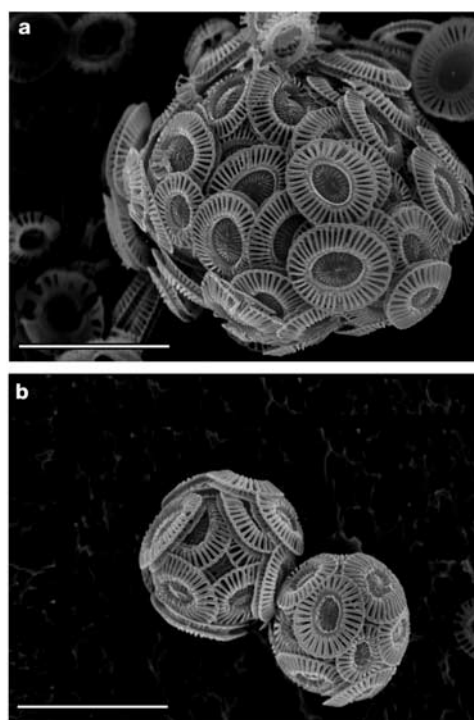


Figure 4 Scanning electron microscopy images of *E. huxleyi* grown under different free zinc and CO_2 concentrations. **a**, 0.3 $pmol\ kg^{-1}$ $[Zn^{2+}]$ and 20 $\mu mol\ kg^{-1}$ $[CO_2]$. **b**, 6 $pmol\ kg^{-1}$ $[Zn^{2+}]$ and 13 $\mu mol\ kg^{-1}$ $[CO_2]$. Scale bars, 5 μm . Note that the effect of CO_2 on calcification is minute compared with that of zinc. See text for details.

ocean by means of atmospheric dust deposition could alter global $CaCO_3$ production on glacial/interglacial timescales. On the one hand, results of a recent study¹⁰ indicate that alleviation of zinc limitation may be accompanied by changes in the natural phytoplankton assemblages in favour of coccolithophores. This would lead to higher coccolithophorid-based biomass and potentially higher total $CaCO_3$ production at the community level. The results of this study, on the other hand, indicate that the amount of $CaCO_3$ that is precipitated in conjunction with a given amount of coccolithophorid biomass produced is expected to decrease with increasing surface-water zinc concentrations. On the basis of this effect, higher atmospheric zinc deposition during glacial periods would correspond to lower $CaCO_3$ production.

If coccolithophorid zinc limitation proves to be a general phenomenon in today's oceans, the observed processes may influence the strength of the ocean's $CaCO_3$ pump. Any change in its intensity directly affects the rain ratio, thereby altering carbon sequestration in the ocean. Long-term variation in atmospheric zinc deposition and its potential effect on biogenic calcification may therefore need to be considered in the context of glacial/interglacial changes in CO_2 partitioning between atmosphere and ocean. □

Methods

Experimental setup

Mono-specific cultures of *E. huxleyi* clone PML B92/11 were grown in triplicates (zinc experiment) or duplicates (iron experiment) at 15 °C in 0.2 μm filtered sea water, originating from the Gulf of Biscay (zinc experiment) and the Antarctic Ocean (iron experiment), at a photon flux density of 180 $\mu mol\ m^{-2}\ s^{-1}$ (supplied from cool white fluorescent bulbs (Philips TLD 36W/54) on a 16/8-h light/dark cycle). Precultures and experimental incubations in dilute batch cultures ensured exponential growth throughout the experiment. The medium was enriched with nitrate and phosphate (64 and 4 $\mu mol\ kg^{-1}$, respectively), vitamin B12 (0.59 $nmol\ kg^{-1}$), biotin (0.2 $nmol\ kg^{-1}$) and 59 $nmol\ kg^{-1}$ of thiamine-HCl (see Fig. 3 legend for details). Final trace metal concentrations were obtained by adding: $FeCl_3$ (1 $\mu mol\ kg^{-1}$) (zinc experiment), $ZnCl_2$ (1 $\mu mol\ kg^{-1}$) (iron experiment), $CuSO_4 \cdot 5H_2O$ (40 $nmol\ kg^{-1}$), $CoCl_2$ (10 $nmol\ kg^{-1}$), $MnCl_2 \cdot 4H_2O$ (450 $nmol\ kg^{-1}$), $(NH_4)_6Mo_7O_{24} \cdot 4H_2O$ (1.43 $nmol\ kg^{-1}$), $Na_2SeO_3 \cdot 5H_2O$

(100 nmol kg⁻¹), KBr (92 μmol kg⁻¹), SrCl₂·6H₂O (13 μmol kg⁻¹), AlCl₃ (100 nmol kg⁻¹), LiCl (70 nmol kg⁻¹), KI (60 nmol kg⁻¹), H₃BO₃ (3.23 μmol kg⁻¹), RbCl (250 nmol kg⁻¹). To achieve defined trace metal conditions for [Zn²⁺] or [Fe(III)]⁺, 1 mmol kg⁻¹ of Na₂EDTA-6H₂O and varying amounts of ZnCl₂ or FeCl₃ were added to the media, which after pH adjustment with NaOH was stored for 24 h in the dark to allow chemical equilibration. In a control experiment at low, widely used culture media [EDTA] of 6 μmol kg⁻¹ the same maximum growth rates were observed, indicating that the comparatively high [EDTA] applied here did not have any detrimental effect on *E. huxleyi*.

Trace metal speciation

[Zn²⁺] and [Fe(III)]⁺ were calculated from total metal concentrations using an equilibrium complexation model. Therefore, conditional stability constants for Fe(III) complexes with chlorine, fluoride, sulphate and the Fe(III) hydroxides (see refs 20, 21 and references therein) were included and, for the sake of consistency, thermodynamic stability constants²² were used for all EDTA complexes with Fe, Cu, Co, Mn, Zn, Ca, Mg, the protonated forms of EDTA, and ZnCO₃ and ZnSO₄ after correction for ionic strength (salinity 34) with ion activity coefficients obtained either by Pitzer modelling²³ or Davies approximation. Light-induced photo-dissociation of Fe-EDTA complexes was accounted for following the model given in ref. 24. On the basis of this model two pH-dependent factors, linearly interpolated to 15 °C, were calculated and corrected for the lower photon-flux density and the light/dark cycle in our experiments (resulting in factors of 1.5 at pH 7.9 and 3 at pH 8.35). The absolute values for [Fe(III)]⁺, however, crucially depend on the selected set of hydrolysis constants, particularly on β₃^{*}, which is so far not well constrained (see Table 2 in ref. 24 and references therein for details). The total zinc and iron concentrations in the natural sea water before nutrient addition were 7 and 1 nmol kg⁻¹, respectively and the [Zn²⁺]/[Zn_{total}] in the zinc experiments was about 1/75,000.

Sampling and measurements

The carbonate system was calculated from pH and total dissolved inorganic carbon using the dissociation constants of ref. 25 as refitted in ref. 26. The pH was measured using the recommendations of ref. 21 and the total dissolved inorganic carbon was measured using a photochemical approach²⁷. To calculate growth rates, cell counts were obtained at the beginning and end of incubations on a Coulter Epics XL-MCL flowcytometer. For measurements of cellular particulate organic carbon (POC) and nitrogen (PON) and total particulate carbon (TPC), subsamples were filtered on precombusted (500 °C) Whatman GF/F filters at the end of the experiments and stored at -25 °C. Before analysis POC filters were fumed for 2 h with concentrated HCl. POC, PON and TPC were analysed on an ANCA-SL 20-20 Europa Scientific mass spectrometer after 2 h of drying the filters at 60 °C. CaCO₃ was calculated by subtracting POC from TPC. CaCO₃ production rates were calculated from growth rates and cellular CaCO₃ contents. Owing to the experimental approach used in this study (see above), these parameters did not change significantly over the course of the experiments.

Received 24 January; accepted 10 May 2004; doi:10.1038/nature02631.

- Petit, J. R. *et al.* Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature* **399**, 429–436 (1999).
- Mahowald, N. *et al.* Dust sources and deposition during the last glacial maximum and current climate: A comparison of model results with paleodata from ice cores and marine sediments. *J. Geophys. Res.* **104**, 15895–15916 (1999).
- Martin, J. H. Glacial-interglacial CO₂ change: the iron hypothesis. *Paleoceanogr.* **5**, 1–13 (1990).
- Morel, F. M. M. *et al.* Zinc and carbon co-limitation of marine phytoplankton. *Nature* **369**, 740–742 (1994).
- Broecker, W. S. & Peng, T.-H. The role of CaCO₃ compensation in the glacial to interglacial atmospheric CO₂ change. *Glob. Biogeochem. Cycles* **1**, 15–29 (1987).
- Milliman, J. D. Production and accumulation of calcium carbonate in the ocean: budget of a nonsteady state. *Glob. Biogeochem. Cycles* **7**, 927–957 (1993).
- Baar de, H. J. W. & Boyd, P. W. in *The Changing Ocean Carbon Cycle* (eds Hanson, R. B., Ducklow, H. W. & Fields, J. G.) 61–140 (Cambridge Univ. Press, Cambridge, 2000).
- Conkright, M. E., Levitus, S. & Boyer, T. P. *World Ocean Atlas 1994*, Vol. 1 *Nutrients* (NOAA Atlas NESDIS 1, US Department of Commerce, Washington DC, 1994).
- Coale, K. H. Effects of iron, manganese, copper, and zinc enrichments on productivity and biomass in the subarctic Pacific. *Limnol. Oceanogr.* **36**, 1851–1864 (1991).
- Crawford, D. W. *et al.* Influence of zinc and iron enrichments on phytoplankton growth in the northeastern subarctic Pacific. *Limnol. Oceanogr.* **48**, 1583–1600 (2003).
- Sunda, W. G. & Huntsman, S. A. Cobalt and zinc interreplacement in marine phytoplankton: Biological and geochemical implications. *Limnol. Oceanogr.* **40**, 1404–1417 (1995).
- Ellwood, M. J. & Van den Berg, C. M. G. Determination of organic complexation of cobalt in seawater by cathodic stripping voltammetry. *Mar. Chem.* **75**, 33–47 (2001).
- Riebesell, U. *et al.* Reduced calcification of marine plankton in response to increased atmospheric CO₂. *Nature* **407**, 364–367 (2000).
- Zondervan, I., Rost, B. & Riebesell, U. Effect of CO₂ concentration on the PIC/POC ratio in the coccolithophore *Emiliania huxleyi* grown under light-limiting conditions and different daylengths. *J. Exp. Mar. Biol. Ecol.* **272**, 55–70 (2002).
- Broecker, W. S. & Peng, T.-H. *Tracers in the Sea* (Eldigio, New York, 1982).
- Bruland, K. W. Complexation of zinc by natural organic ligands in the central North Pacific. *Limnol. Oceanogr.* **34**, 269–285 (1989).
- Krennig, K. & Streu, P. The behaviour of dissolved Cd, Co, Zn and Pb in North Atlantic near-surface waters (30°N/60°W–60°N/2°W). *Deep-Sea Res.* **1** **48**, 2541–2567 (2001).
- Lohan, M. C., Statham, P. J. & Crawford, D. W. Total dissolved zinc in the upper water column of the subarctic North East Pacific. *Deep-Sea Res.* **II** **49**, 5793–5808 (2002).
- Duce, R. *et al.* The atmospheric input of trace species to the world ocean. *Glob. Biogeochem. Cycles* **5**, 193–259 (1991).
- Liu, X. & Millero, F. J. The solubility of iron in seawater. *Mar. Chem.* **77**, 43–54 (2002).

- DOE. *Handbook of Methods for the Analysis of the Various Parameters of the Carbon Dioxide System in Seawater* (eds Dickson, A. G. & Goyet, C.) Version 2.1 ORNL/CDIAC-74 (<http://andrew.ucsd.edu/co2qc/handbook.html>) (1994).
- Pettit, L. D. & Powell, K. J. *IUPAC Stability Constants Database* (IUPAC and Academic Software, Otley, 2001).
- Millero, F. J. & Pierrot, D. A chemical equilibrium model for natural waters. *Aquat. Geochem.* **4**, 153–199 (1998).
- Sunda, W. & Huntsman, S. Effect of pH, light, and temperature on Fe-EDTA chelation and Fe hydrolysis in seawater. *Mar. Chem.* **84**, 35–47 (2003).
- Mehrbach, C., Culbertson, C. H., Hawley, J. E. & Pytkowicz, R. N. Measurement of the apparent dissociation constants of carbonic acid in seawater at atmospheric pressure. *Limnol. Oceanogr.* **18**, 897–907 (1973).
- Dickson, A. G. & Millero, F. J. A comparison of the equilibrium constants for the dissociation of carbonic acid in seawater media. *Deep-Sea Res.* **34**, 1733–1743 (1987).
- Stoll, M. H. C., Bakker, K., Nobbe, G. H. & Haese, R. R. Continuous flow analysis of dissolved inorganic carbon content in seawater. *Anal. Chem.* **73**, 4111–4116 (2001).
- Candelone, J.-P., Hong, S., Pellone, C. & Bontouron, C. F. Post-Industrial Revolution changes in large-scale atmospheric pollution of the northern hemisphere by heavy metals as documented in central Greenland snow and ice. *J. Geophys. Res.* **100**, 16605–16616 (1995).
- Hong, S., Candelone, J.-P. & Bontouron, C. F. Changes in zinc and cadmium concentrations in Greenland ice during the past 7760 years. *Atmos. Environ.* **31**, 2235–2242 (1997).
- Hong, S., Candelone, J.-P., Turetta, C. & Bontouron, C. F. Changes in natural lead, copper, zinc and cadmium concentrations in central Greenland ice from 8250–149,100 years ago: their association with climate changes and resultant variations of dominant source contributions. *Earth Planet. Sci. Lett.* **143**, 233–244 (1996).

Acknowledgements We thank A. Terbrüggen, K.-U. Richter and B. van der Wagt for laboratory assistance, and M. Lohan, K. W. Bruland and R. E. Zeebe for discussions during the preparation of this manuscript. This work was partly funded by the German Research Foundation (DFG).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to K.G.S. (kschulz@ifm-geomar.de).

Aggression by polyembryonic wasp soldiers correlates with kinship but not resource competition

David Giron¹, Derek W. Dunn², Ian C. W. Hardy² & Michael R. Strand¹

¹Department of Entomology, University of Georgia, Athens, Georgia 30602, USA

²School of Biosciences, University of Nottingham, Sutton Bonington Campus, Loughborough LE12 5RD, UK

Kin selection theory predicts that individuals will show less aggression and more altruism towards relatives^{1,2}. However, recent theoretical developments suggest that with limited dispersal, competition between relatives can override the effects of relatedness^{3–9}. The predicted and opposing influences of relatedness and competition are difficult to approach experimentally because conditions that increase average relatedness among individuals also tend to increase competition. Polyembryonic wasps in the family Encyrtidae are parasites whose eggs undergo clonal division to produce large broods¹⁰. These insects have also evolved a caste system: some embryos in a clone develop into reproductive larvae that mature into adults, whereas others develop into sterile soldier larvae that defend siblings from competitors^{11–14}. In a brood from a single egg, reproductive altruism by soldiers reflects clone-level allocation to defence at the cost of reproduction, with no conflict between individuals. When multiple eggs are laid into a host, inter-clone conflicts of interest arise. Here we report that soldier aggression in *Copidosoma floridanum* is inversely related to the genetic relatedness of competitors but shows no correlation with the level of resource competition.

Polyembryonic encyrtids are small (1 mm) parasitoid wasps that

oviposit into moth (host) eggs^{10,11}. After parasitism the host egg hatches and the larva develops to its final (fifth) instar (Fig. 1). During this period, the wasp egg proliferates clonally, producing multiple embryos in an assemblage called a polymorula. Several species, including *C. floridanum*, produce thousands of embryos per egg. Some embryos develop during the host's early instars into obligately sterile soldier (precocious) larvae with fighting mandibles and an elongate body^{10–12}. The remaining embryos develop during the host's final instar into reproductive larvae with tiny mandibles and rounded bodies. These reproductive larvae consume the host, pupate and emerge as adult wasps. Soldier larvae always die after the reproductive larvae consume the host¹⁰.

Hamilton's rule specifies that altruism is favoured when $rb - c > 0$, where c is the fitness cost to the altruist, b is the fitness benefit to the recipient and r is their genetic relatedness^{1,2}. Conditions favouring the most extreme form of altruism, that is, sterile castes, include high relatedness and/or factors that make $b > c$. Key elements in the evolution of sterile soldier larvae by polyembryonic wasps are thus likely to be: (1) clonal development in a confined space (high r) and (2) the benefits, b , of defending a nutrient-rich but limiting resource (the host) from competitors relative to the costs, c , in lost reproduction^{10,11}. Previous studies indicate that soldiers attack and frequently sacrifice themselves in defence of their clone-mates against other species of parasitoids^{13,14}. In laboratory experiments, 96% of females with no previous oviposition experience ($N = 50$) oviposited into hosts that were parasitized by another wasp (superparasitism). Females superparasitized hosts irrespective of when the first female oviposited or the relatedness of the resident clone. Screening individual wasps from broods (56 broods with 20 same sex wasps per brood) collected in Georgia, using polymorphic allozymes such as glucosephosphate isomerase, further indicated that some broods contain progeny of at least two genotypes (D.G., K. G. Ross and M.R.S., unpublished results). Given that females also often lay two eggs per host (see below), these data collectively indicate that: (1) different kin classes of *C. floridanum* frequently compete within the same host, and (2) relatedness of potential competitors ranges from full siblings to distant relatives. This provides an opportunity for experimentally separating the effects of relatedness and competition under conditions of limited dispersal while using soldier behaviour to quantify conflicts of interest.

Like most Hymenoptera *C. floridanum* is haplodiploid, with unfertilized eggs producing males and fertilized eggs producing

females. Mothers produce all-male or all-female broods by laying one egg per host, and mixed broods by laying two eggs (one male and one female)¹⁵. Almost 1,400 wasps per host emerge from each brood type^{16,17}. A developmental asymmetry, however, exists between males and females: female broods produce about 30–50 soldiers by the host's fourth instar and male broods produce almost none^{15–18}. Correspondingly, soldiers in mixed broods are almost exclusively female. Female soldiers have been shown to attack male embryos which results in mixed broods having strongly female-biased sex ratios^{16,19}. The resolution of conflict in mixed broods in favour of females could be due to relatedness asymmetries, because soldiers are predicted to value reproductive, clone-mate sisters ($r = 1$) more than brothers ($r \sim 0.25$), particularly as only a few of the latter are needed for sib mating¹⁶. The indirect pay-off (inclusive fitness) to soldier larvae through a brother's post-dispersal matings may, however, provide some check on siblicidal behaviour¹⁹.

If soldiers are able to assess their relatedness to other individuals in the host, we might expect the aggressive behaviour of soldiers to be negatively correlated with relatedness to a competing clone. The highest levels of inter-clone aggression should thus be observed when unrelated mothers lay eggs into the same host. The completely local level of competition for limiting host resources could, however, override effects of relatedness^{3–9}, in which case soldier behaviour towards other clones is predicted to correlate with the severity of resource competition. This will increase with the number of competitors and/or reductions in host size, irrespective of relatedness. An additional consideration is that female clones might benefit from the presence of future mating partners if mating opportunities after dispersal are scarce. If so, soldiers might exhibit lower levels of aggression towards non-relative males than females.

To evaluate the importance of relatedness asymmetries, resource competition and sex of the competitor on soldier aggression, we conducted two complementary experiments. In the first experiment, we injected a labelled polymorula that was either a full sibling (brother or sister) or a non-relative (male or female) into a host containing an all-female brood (see Methods). Note that introducing different kin classes into hosts containing a female brood varied relatedness (and sex) of the potential competitors. However, the intensity of resource competition among treatments was approximately the same because each host was almost identical in size and host resources for the resident clone would be similarly reduced by injection of a second polymorula regardless of its relatedness. Hosts containing a female brood were also injected with a labelled tissue (gonad) collected from a non-parasitized host larva. This tissue would have no impact on relatedness, resource competition or future mating opportunities, and controlled for the possibility that soldiers attack any foreign entity introduced into the host. In all treatments, attack of the target by resident soldiers was scored as: (1) the proportion of hosts from each treatment that contained at least one soldier with fluorescent tracer in its gut, and (2) the mean proportion of soldiers per host with tracer in their gut. Both measures revealed a strong association between soldier aggression and relatedness of the competitor (Table 1). Attack rates were highest towards non-relative female or male clones, intermediate towards brothers and lowest towards full sisters or host tissue. Resource competition did not seem to affect attack rate because a sister clone would reduce available host resources whereas host tissue (gonad) would not, but resident soldiers rarely attacked either. In contrast, soldiers attacked non-relative clones much more frequently than sisters even though all introduced competitors would similarly reduce host resources for the resident clone. Soldier aggression also did not correlate with the sex of the competitor and future mating opportunities, given that soldiers attacked males as frequently as non-relative females. Direct observation of soldier larvae *in vitro* corroborated our *in vivo* results by showing that non-relative clones and brothers were attacked significantly more often

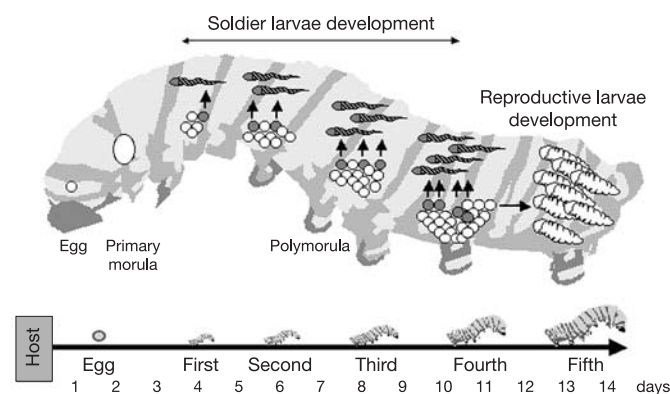


Figure 1 Life cycle of *C. floridanum* in its host *T. ni*. The schematic shows a host larva and the developmental stages of *C. floridanum*. Below the schematic are the host's life stages in relation to time (days). After parasitism, the host egg hatches and the larva develops to a fifth instar over 14 days. The *C. floridanum* egg develops initially into a primary morula. The primary morula then clonally proliferates to form a polymorula of more than 1,000 embryos during the host first to fourth instar period. Soldier larvae eclose during this period. Reproductive larvae eclose in the host fifth instar.

Table 1 Proportion of labelled competitors attacked by soldiers *in vivo*

Relatedness of the competitor (N)	Resource competition	Future mating opportunities	Proportion of hosts with labelled soldiers	Mean proportion of soldiers labelled per host
Sister (40)	Yes	No	0.12	0.001 ^a
Brother (31)	Yes	Yes	0.52	0.09 ^b
Unrelated female, Georgia (40)	Yes	No	0.68	0.16 ^{b, c}
Unrelated female, Wisconsin (30)	Yes	No	0.76	0.21 ^c
Unrelated male, Wisconsin (31)	Yes	Yes	0.77	0.24 ^c
Gonad (control, 31)	No	No	0.07	0.001 ^a

The proportion of hosts containing at least one labelled soldier varied significantly with relatedness of the competitor, resource competition and future mating opportunities (Full model, $G^2 = 75.43$; d.f. = 7; $P < 0.0001$). Calculation of likelihood ratios for each effect revealed that relatedness of the competitor was of greatest significance ($G^2 = 48.33$; d.f. = 5; $P < 0.0001$). Attack rates towards all sisters did not differ significantly from controls (host gonads) ($G^2 = 0.75$; d.f. = 1; $P > 0.40$), suggesting that resource competition did not affect attack behaviour. Soldier attack rates towards males and unrelated females did also not significantly differ ($G^2 = 2.75$; d.f. = 1; $P > 0.25$), indicating that soldier aggression was unaffected by the sex of the competitor and potential future mating opportunities. A similar association between relatedness of the competitor and soldier aggression was obtained when the mean proportion of labelled soldiers per host was determined. A mean ($\pm$ s.e.m.) of 56.6 ± 16.2 soldiers were present per host with no significant differences between treatments ($F_{5,203} = 1.72$; $P > 0.50$). Almost no soldiers were labelled when challenged by sisters or host gonads, whereas an increasing proportion of the soldiers per host contained label in their gut when the competitor was a brother or less-related clone ($F_{5,203} = 27.6$; $P < 0.0001$; arcsin transformed data; means with the same superscript letter (a, b, c) were not significantly different (Tukey–Kramer method)).

than sisters ($P < 0.0001$) (Table 2). Combined, these results clearly indicate that *C. floridanum* soldiers differentially attack competitors as a function of relatedness. Recent studies further reveal that soldiers distinguish kin from non-kin by cues from the extraembryonic membrane that surrounds each individual²⁰. This membrane is also essential for protecting *C. floridanum* from the host's immune response, suggesting that the cues used in kin recognition may be maintained in part by selection for resistance against the host.

In our second experiment, we simultaneously assessed the effect of resource competition and relatedness on soldier aggression by comparing attacks towards competitors of different relatedness in large versus small hosts. Large hosts were naturally parasitized larvae (unstarved) as used in experiment 1, whereas small hosts were created by depriving parasitized larvae of food for 48 h. Hosts always survived transient starvation but their average maximum size (286 mg) was reduced by almost 50% relative to unstarved hosts (507 mg) (see Methods). Host starvation clearly intensified resource competition among progeny of the resident clone, because the number of adult progeny produced in starved hosts (485 ± 74 (mean $\pm$ s.e.m.), $N = 20$) was significantly smaller than in unstarved hosts (mean = $1,206 \pm 117$, $N = 20$; $P < 0.0001$ (*t*-test)) (see Methods). By extension, the intensity of resource competition after introduction of a second polymorula would likewise be much higher between kin classes in a starved host compared with an unstarved host. We therefore injected labelled sister, brother, non-relative female (Wisconsin) or non-relative male polymorulae into starved or unstarved hosts containing a Georgia female clone ($N = 30$ for each treatment). When we compared the outcomes among only starved hosts, we found that soldier aggression remained strongly associated with the relatedness of the competitor ($G^2 = 53.87$; d.f. = 4; $P < 0.0001$). This indicates that soldiers differentially attacked competitors as a function of relatedness even when host resources were more limited than in experiment 1. Moreover, we found no differences in soldier attack rates towards sisters (unstarved, 0.18; starved, 0.12; $P > 0.5$), brothers (unstarved, 0.52; starved, 0.5; $P > 0.8$), non-relative females (unstarved, 0.76; starved, 0.80; $P > 0.9$) or non-relative males (unstarved, 0.77; starved, 0.76; $P > 0.9$) when we compared outcomes in starved hosts versus unstarved hosts. This indicated

that levels of soldier aggression are negatively correlated with relatedness of the potential competitor, but do not change with the severity of resource competition.

Hamilton originally proposed that the evolution of altruism could arise through either direct kin recognition or limited dispersal (population viscosity) which would have the effect of elevating local relatedness sufficiently to favour generalized altruism towards neighbours^{1,2}. Our results with *C. floridanum* indicate that altruism extends beyond clone-mates to close relatives like non-identical sisters. This response is unlikely to reflect generalized altruism due to high neighbour–neighbour relatedness because females do not benignly share resources with brothers or non-relatives. Recent theory further suggests that altruism is unlikely to evolve in purely viscous populations because viscosity increases neighbour–neighbour competition for resources as much as it elevates relatedness^{3–9}. Potential benefits of sharing a host with future mating partners also seem to be relatively unimportant in *C. floridanum* because soldiers usually attack males. Instead, our data support the alternative conclusion that altruism in *C. floridanum* operates by means of kin recognition^{1,2}. Theory suggests that the evolution of kin-recognition based altruism may be suppressed in viscous populations unless periodic dispersal (alternating viscosity) occurs, which allows the beneficiaries of altruism to compete for resources against non-relatives^{3,4,6}. Within the Georgia field population that was used in this study, the frequency of homozygous broods for the polymorphic allozyme glucosephosphate isomerase depart significantly from Hardy–Weinberg expectations ($P < 0.002$; Fisher's exact test) suggesting a high level of inbreeding (D.G., K. G. Ross and M.R.S., unpublished results). However, adult dispersal and subsequent resource competition between non-relatives is also likely because winged adult females must disperse to find new hosts and winged adult males are likewise able to disperse and can potentially mate with females from other hosts^{15,16}. Our observations of sister–sister ($r \sim 0.75$) altruism but sister–brother ($r \sim 0.25$) conflict within hosts further suggest that opportunities exist for surviving males from mixed broods to disperse and mate with non-sisters¹⁶. In contrast, a comparative study of fig wasps suggests that local competition between males for mates is severe enough to obviate the benefits of kin selection⁸.

The critical interplay between relatedness and competition in the evolution of altruism is potentially best understood in organisms that develop in non-standard ways, like clonal species that form assemblages of potentially competing relatives and non-relatives²¹. Like *C. floridanum*, several other clonally developing animals show a strong ability to recognize close relatives^{22–24}. In contrast, examples of one clone exploiting another when clones mix are rare, and may reflect either the inability of some species to suppress conflict, or that a sufficiently large benefit to cost ratio ($b > c$) favours cooperation in the absence of high relatedness^{25,26}. In *C. floridanum*, soldier attack clearly depends on relatedness, even under severe resource competition. □

Table 2 Proportion of competitors attacked by soldiers *in vitro*

Competitor	N	Proportion of competitors attacked
Sister	30	0.16
Brother	30	0.57
Unrelated female (Georgia)	30	0.60
Unrelated female (Wisconsin)	29	0.76
Unrelated male (Wisconsin)	31	0.71

Sister clones are attacked less than brothers or non-relative clones ($G^2 = 24.63$; d.f. = 1; $P < 0.0001$).

Methods

Two populations of *C. floridanum* were established in the laboratory from field-collected broods obtained in the southern (Georgia) and northern (Wisconsin) US. Each culture was maintained separately as large randomly mating populations on the host *Trichoplusia ni*. Host larvae were reared on an artificial diet at 27 °C and in a 16 h light/8 h dark photoperiod as previously described²⁷. Experimental conditions were manipulated to favour production of all-male and all-female broods. Ovipositions were observed to determine that wasps laid one male or female egg to produce all-female or all-male broods¹⁷. Identifying the composition of broods in advance was possible, because females exhibit different behaviours when laying a fertilized or an unfertilized egg¹⁴. Hosts containing full sisters were produced by mating clonal females from one brood with a male from another. Hosts containing brothers were produced by allowing unmated females from the same brood to oviposit.

Male or female polymorulae containing on average 500 embryos were collected from fourth instar hosts and labelled with carboxyfluorescein diacetate succinimidyl ester (CFSE) using previously established methods^{14,16}. These previous studies confirmed that soldiers that attack CFSE-labelled tissues ingest the label which is then clearly visible when larvae are examined by epifluorescent microscopy. In experiment 1, CFSE-labelled sister, brother, non-relative Georgia female, non-relative Wisconsin female and non-relative Wisconsin male polymorulae were injected into fifth instar hosts containing a female brood from the Georgia population using a glass needle mounted on a micromanipulator. As a control, testes from fourth instar *T. ni*, which are similar in size to a polymorula, were labelled and individually injected. Hosts were then dissected 24 h after injection and the number of soldiers with label in their gut were determined. *In vitro* assays were conducted in 1-ml culture wells containing TC-100 medium (Sigma)^{14,16}. Soldiers and polymorulae from relative and non-relative broods were collected from fourth instar hosts immediately before the experiment. One soldier larva was placed in the culture well with a full sister, brother, non-relative female or non-relative male polymorula. The proportion of each type of polymorula that was attacked during a 2-h bioassay period was then recorded. An attack was defined as the larva gripping a polygerm with its mandibles for more than 1 min. When this occurred, consumption of tissue by the soldier was readily visible.

In experiment 2, hosts containing female broods were starved by removing them on the first day of the fifth instar from the food source to cups containing moist cotton wool. Each larva was returned to food 24 or 48 h after starvation and reared until completion of *C. floridanum* development. The average maximum size of hosts treated in this manner declined linearly from 507 ± 56 mg (mean ± s.e.m., *N* = 20) for unstarved controls to 398 ± 42 mg (*N* = 20) and 286 ± 38 mg (*N* = 20) for hosts starved for 24 or 48 h, respectively. Previous studies indicated that the total number of wasp progeny produced per polymorula is established at the end of the host fourth instar^{11,18}. Thus, any reduction in the number of progeny produced in starved, fifth instar hosts is due to mortality caused by increased competition for more limited host resources rather than a change in the development of the polymorula itself. To assess the effect of host starvation on the resident clone, parasitized hosts containing a female brood were starved for 48 h, returned to food and then reared until completion of development. The total number of wasp progeny per host was then counted. For comparisons of soldier aggression in starved and unstarved hosts, one cohort of hosts was starved for 48 h, returned to food and then injected 24 h later with a CFSE-labelled sister, brother, non-relative female (Wisconsin) or non-relative male polymorula. The second, unstarved cohort was treated identically except that hosts were continuously provided with food. Hosts in both cohorts were dissected 24 h after injection and the number of hosts containing soldiers with label in their gut was determined as described in experiment 1.

The association between soldier aggression, relatedness, resource competition or gender of the competitor were analysed by likelihood ratio chi-square tests using the JMP, v.3.0 statistical package.

Received 30 March; accepted 7 June 2004; doi:10.1038/nature02721.

- Hamilton, W. D. The evolution of altruistic behaviour. *Am. Nat.* **97**, 354–356 (1963).
- Hamilton, W. D. The genetical evolution of social behaviour, I & II. *J. Theor. Biol.* **7**, 1–52 (1964).
- Wilson, D. S., Pollock, G. B. & Dugatkin, L. A. Can altruism evolve in a purely viscous population? *Evol. Ecol.* **6**, 331–341 (1992).
- Taylor, P. D. Altruism in viscous populations—an inclusive fitness model. *Evol. Ecol.* **6**, 352–356 (1992).
- Taylor, P. D. Inclusive fitness in a homogeneous environment. *Proc. R. Soc. Lond. B* **249**, 299–302 (1992).
- Queller, D. C. Genetic relatedness in viscous populations. *Evol. Ecol.* **8**, 70–73 (1994).
- Grafen, A. in *Behavioural Ecology: An Evolutionary Approach* (eds Krebs, J. R. & Davies, N. B.) 62–84 (Blackwell Scientific, Oxford, 1984).
- West, S. A., Murray, M. G., Machado, C. A., Griffin, A. S. & Herre, E. A. Testing Hamilton's rule with competition between relatives. *Nature* **409**, 510–513 (2001).
- West, S. A., Penn, I. & Griffin, A. S. Cooperation and competition between relatives. *Science* **296**, 72–75 (2002).
- Strand, M. R. in *Encyclopedia of Insects* (eds Cardé, R. & Resch, V.) 928–932 (Academic, San Diego, 2003).
- Strand, M. R. & Grbic, M. The development and evolution of polyembryonic insects. *Curr. Top. Dev. Biol.* **35**, 121–160 (1997).
- Donnell, D., Corley, L. S., Chen, G. & Strand, M. R. Caste determination in a polyembryonic wasp involves inheritance of germ cells. *Proc. Natl Acad. Sci. USA* (in the press).
- Cruz, Y. P. A sterile defender morph in a polyembryonic hymenopterous parasite. *Nature* **294**, 446–447 (1981).
- Harvey, J. A., Corley, L. S. & Strand, M. R. Competition induces adaptive shifts in caste ratios of a polyembryonic wasp. *Nature* **406**, 183–186 (2000).
- Strand, M. R. Oviposition behavior and progeny allocation by the polyembryonic wasp *Copidosoma floridanum*. *J. Insect Behav.* **2**, 355–369 (1989).
- Grbic, M., Ode, P. J. & Strand, M. R. Sibling rivalry and brood sex ratios in polyembryonic wasps. *Nature* **360**, 254–256 (1992).

- Ode, P. J. & Strand, M. R. Progeny and sex allocation decisions of the polyembryonic wasp *Copidosoma floridanum*. *J. Anim. Ecol.* **64**, 213–224 (1995).
- Grbic, M., Nagy, L., Carroll, S. B. & Strand, M. R. Development and pattern formation in the polyembryonic wasp, *Copidosoma floridanum*. *Development* **122**, 795–804 (1996).
- Ode, P. J. & Hunter, M. S. in *Sex ratios: Concepts and Research Methods* (ed. Hardy, I. C. W.) 218–234 (Cambridge Univ. Press, Cambridge, 2002).
- Giron, D. & Strand, M. R. Host resistance and the evolution of kin recognition in polyembryonic wasps. *Proc. R. Soc. Lond. (Suppl.) Biol. Lett.* published online 17 June 2004 (DOI:10.1098/rsbl/2004.0205).
- Queller, D. C. Relatedness and the fraternal major transitions. *Phil. Trans. R. Soc. Lond. B* **355**, 1647–1655 (2000).
- Crozier, R. H. Genetic clonal recognition abilities in marine invertebrates must be maintained by selection for something else. *Evolution* **40**, 1100–1101 (1986).
- Buss, L. W. *The Evolution of Individuality* (Princeton Univ. Press, Princeton, 1987).
- Grosberg, R. K. The evolution of allorecognition specificity in clonal invertebrates. *Q. Rev. Biol.* **63**, 377–412 (1988).
- Strassmann, J. E., Zhu, Y. & Queller, D. C. Altruism and social cheating in the social amoeba *Dictyostelium discoideum*. *Nature* **408**, 965–967 (2000).
- Abbot, P., Withgott, J. H. & Moran, N. A. Genetic conflict and conditional altruism in social aphid colonies. *Proc. Natl Acad. Sci. USA* **98**, 12068–12071 (2001).
- Strand, M. R. Development of the polyembryonic parasitoid *Copidosoma floridanum* in *Trichoplusia ni*. *Entomol. Exp. Appl.* **50**, 37–46 (1989).

Acknowledgements This work was supported in part by the Natural Environment Research Council (UK), the National Science Foundation (US), the University of Georgia Experiment Station, and the Conseil General de la Region (France).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.R.S. (mrstrand@bugs.ent.uga.edu).

High mutation rate and predominance of insertions in the *Caenorhabditis elegans* nuclear genome

Dee R. Denver¹, Krystalynne Morris², Michael Lynch¹, & W. Kelley Thomas²

¹Department of Biology, Indiana University, Bloomington, Indiana 47405, USA

²Hubbard Center for Genome Studies, University of New Hampshire, Durham, New Hampshire 03824, USA

Mutations have pivotal functions in the onset of genetic diseases and are the fundamental substrate for evolution. However, present estimates of the spontaneous mutation rate and spectrum are derived from indirect and biased measurements. For instance, mutation rate estimates for *Caenorhabditis elegans* are extrapolated from observations on a few genetic loci with visible phenotypes and vary over an order of magnitude¹. Alternative approaches in mammals, relying on phylogenetic comparisons of pseudogene loci² and fourfold degenerate codon positions³, suffer from uncertainties in the actual number of generations separating the compared species and the inability to exclude biases associated with natural selection. Here we provide a direct and unbiased estimate of the nuclear mutation rate and its molecular spectrum with a set of *C. elegans* mutation-accumulation lines that reveal a mutation rate about tenfold higher than previous indirect estimates and an excess of insertions over deletions. Because deletions dominate patterns of *C. elegans* pseudogene variation^{4,5}, our observations indicate that natural selection might be significant in promoting small genome size, and challenge the prevalent assumption that pseudogene divergence accurately reflects the spontaneous mutation spectrum.

Table 1 Insertion and deletion mutations in the MA lines

Chr.	Chr. Pos.	Cos./YAC	Mut.	Context	Line(s)	Cod.
I	8,738,889	C36B1	+A	TCGTT → TCGATT	23	IN
I	8,739,217	C36B1	− C	TTCTT → TTCT	73	IN
I	8,952,991	F53B6	+C	CCTTT → CCTCTT	39	IN
I	12,826,513	B0019	+G	CCGTT → CCGGTT	12	IN
II	3,170,985	Y25C1A	+G	TAGCG → TAGGCG	95	IN
II	5,629,513	C17G10	+C	AGCAA → AGCCAA	13	IG
II	5,924,746	F59G1	+G	TGCAG → TGGCAG	2	EX
II	5,924,746	F59G1	+C	TGCAG → TGCCAG	14	EX
II	13,314,289	Y48C3A	+CCC	AGCCCCG → AGCCCCCGC	80	EX
II	13,314,321	Y48C3A	+GAT	TGGATCA → TGGATGATCA	80	EX
II	13,314,343	Y48C3A	+GA	CGAGCA → CGAGAGCA	80	EX
III	12,353,930	Y75B8A	− 11 bp	(T ₇ GGTC) ₂ → (T ₇ GGTC) ₁	90	IN
IV	~1,564,830*	Y41D4B	+ ~ 500 bp	ND	46	IN
IV	17,277,013	F26D10	+T	ACTTG → ACTTTG	89	EX
V	11,074,113	T04C12	+G	ATGGCG → ATGGGCG	39	IG
X	10,021,926	F19C6	− 66 bp	(17 bp) (49 bp) (17 bp) → (17 bp)	29	IN
X	10,172,717	C07A4	− A	GC(A) ₅ TA → GC(A) ₄ TA	2	EX

Chr. indicates the chromosome on which the mutation was found, and Chr. Pos. refers to the chromosomal position of the mutation. Cos./YAC refers to the sequenced *C. elegans* cosmid or yeast artificial chromosome in which the mutation was found. Mut. refers to the observed mutation. Line(s) indicates the MA line number(s) where the mutation was detected. Cod. refers to the type of coding sequence in which the mutation was detected: EX is exon, IG is intergenic, and IN is intron.

*A large insertion mutation whose sequence was not determined (Supplementary Fig. S2).

To obtain a direct estimate of the spontaneous mutation rate and spectrum, we sequenced more than 4 megabases of nuclear DNA from a set of *C. elegans* mutation-accumulation (MA) lines. The MA lines were simultaneously initiated from a single N2 (laboratory strain) ancestor and propagated in a benign environment across generations by single-progeny descent, ensuring that all but the most deleterious mutations accumulated over time in an effectively neutral fashion, and resulting in marked declines⁶ in fitness. We sequenced 29,561 base pairs (bp) from each of 72 MA lines at 280 generations, 14,550 bp from each of 68 MA lines at 353 generations, and 18,718 bp from each of 58 MA lines at 396 generations (Supplementary Table S1). All mutations were visually confirmed on electropherogram data from both strands of directly sequenced polymerase chain reaction (PCR) products, ensuring a very low error rate (see Methods). The same loci were sequenced from a set of natural isolates of *C. elegans* for a direct comparison of the mutation spectrum in the MA lines with natural substitution patterns⁷.

We detected 30 mutations (Tables 1 and 2) in the MA lines, yielding a mutation rate estimate of 2.1×10^{-8} mutations per site per generation (standard error of the mean, s.e.m. = $\pm 6.7 \times 10^{-9}$) or, based on an average generation time of four days, 2.0×10^{-6} mutations per site per year ($\pm 6.1 \times 10^{-7}$). The total haploid genomic mutation rate (U_1) is ~ 2.1 mutations per genome per generation, given that the size of the *C. elegans* genome is $\sim 10^8$ bp. The distribution of mutations in each MA line and each locus were very close to Poisson expectations (see Methods). The nuclear mutation rate estimate reported here is about one order of magnitude lower than the mitochondrial rate detected in the same *C. elegans* MA lines (1.6×10^{-7} mutations per site per generation)⁸ but is nearly one order of magnitude higher than the previous average gene-specific nuclear rate estimate for *C. elegans*¹. This observation is consistent with a recent study showing laboratory-based mutation rate estimates about tenfold higher than phylogenetic estimates for *Escherichia coli* and *Salmonella enterica*⁹.

More than half of the mutations (17 of 30) observed in the MA lines were insertion–deletion (indel) mutations (Table 1); among the indels, insertions were dominant (13 of 17). The latter observation differs markedly from patterns of pseudogene evolution in natural populations of both *Caenorhabditis* and *Drosophila*, in which deletions outnumber insertions^{4,5,10}. Among *C. elegans* transposon pseudogenes, for instance, deletions are 2.8-fold more frequent than insertions⁵. This apparent bias towards deletions inferred from patterns of pseudogene evolution is often invoked as a mutation-based mechanism responsible for the maintenance of small genome size in metazoans such as *Drosophila* and *C. elegans*^{10,11}. However, several studies indicate that pseudogenes might

be subject to selective constraints: the makorin1 p1 mouse pseudogene is transcribed and regulates the messenger RNA stability of its homologous coding gene¹², and transcriptional activity is reported for $\sim 20\%$ of *Arabidopsis* sequences annotated as pseudogenes¹³. Furthermore, many pseudogenes in a variety of species display codon bias, retention of reading frame, and very low ratios of nonsynonymous to synonymous substitution rates (K_a/K_s) (ref. 14). The predominance of insertions observed in the MA lines coupled with the bias towards deletions in *C. elegans* pseudogenes indicates that long-term pseudogene evolution in *C. elegans* might not accurately reflect the baseline spontaneous mutation spectrum and that natural selection might favour deletions and/or prevent insertions in otherwise neutral regions of the *C. elegans* genome. If this interpretation is correct, the evolution of the compact *C. elegans* genome might be driven largely by selective rather than mutational forces¹⁵.

Although on the surface it might seem unlikely that natural selection can act on small insertions in presumably non-functional regions of the genome, selection is much more efficient in invertebrate species with large effective population sizes, such as *C. elegans*, than in those with relatively small effective population sizes, such as vertebrates, as reflected in many aspects of genome evolution, including numbers and sizes of introns and the abundances of mobile elements¹⁶. Furthermore, natural selection prevents the accumulation of other types of mutation with seemingly mild effects, such as unpreferred base substitutions at silent sites in highly transcribed genes ('codon bias')¹⁷ and spurious transcription-factor binding sites¹⁸. In species with very small effective

Table 2 Base substitution mutations in the MA lines

Chr.	Chr. Pos.	Cos./YAC	Mut.	Context	Line(s)	Cod.
I	6,816,148	K02F2	C → T	TACTA → TATTA	64, 78	IN
I	8,738,859	C36B1	A → G	GGATA → GGGTA	90	IN
II	6,537,060	C56E6	G → A	AAGCC → AAACC	97	IG
II	10,851,146	M106	G → C	ACGTC → ACCTC	13	IG
III	1,933,779	F53A3	G → C	GAGAT → GACAT	77	IN
III	1,933,788	F53A3	C → G	AACCT → AAGCT	77	IN
III	12,353,902	Y75B8A	T → C	AATTT → AACTT	26	IN
IV	8,001,467	Y42H9B	G → A	TAGAG → TAAAG	67	EX (N: S → F)
IV	10,332,063	K11E8	C → T	GCCGT → GCTGT	61	IN
V	11,724,218	B0240	C → A	GTCAG → GTCAA	84	IG
X	10,173,063	C07A4	G → A	TCGGC → TCAGC	94	EX (S)
X	10,609,426	F59F5	A → T	CAACA → CATCA	98	EX (N: K → R)

Chr. indicates the chromosome on which the mutation was found, and Chr. Pos. refers to the chromosomal position. Cos./YAC refers to the sequenced *C. elegans* cosmid or yeast artificial chromosome in which the mutation was found. Mut. refers to the observed mutation. Line(s) indicates the MA line number(s) where the mutation was detected. Cod. refers to the coding context of the sequence in which the mutation was found: EX is exon, IG is intergenic, and IN is intron. For exon mutations, S indicates a synonymous mutation and N indicates a nonsynonymous mutation (amino acid change is indicated after N).

population sizes, such as humans, in which natural selection is less efficient, pseudogene variation might more accurately reflect the spontaneous mutation spectrum.

Our observations are also based on organisms propagated in the laboratory, whereas previous comparative phylogenetic approaches to estimating the mutation spectrum, although indirect, are derived from sequences evolving in nature. Although we cannot exclude the possibility that mutation processes might differ between laboratory-reared organisms and those occurring in nature, observations in prokaryotes indicate that certain mutation processes (the bias towards transitions over transversions, for instance) do not significantly differ between bacteria raised in the laboratory and those evolving in nature⁹. Furthermore, the rate of sex-linked lethal mutations is highly similar in wild-caught and laboratory populations of *Drosophila*¹⁹.

Five frameshift-causing indel mutations were observed in exon sequence, each resulting in a premature stop codon (Table 1). Two of these indels occurred independently in MA2 and MA14 at the same precise location in the final exon of the *vps-35* gene (F59G1.3). Length variation was also observed at this precise location in the *C. elegans* natural isolates⁷, indicating that this might be a hotspot for insertion mutations. Three short insertions were observed within a 60-bp segment of the fourth exon of the *mac-1* gene (Y48C3A.7) in MA80 (Supplementary Fig. S1), an insertion of a single base pair was detected in the second exon of the *hsp-1* gene, and a single base-pair deletion occurred in the third exon of an unnamed gene (C07A4.2). Although the functional consequences of these specific mutations remain uncertain, *C. elegans* RNA-mediated interference (RNAi) studies²⁰ provide some insights: for the *mac-1* and C07A4.2 genes RNAi treatment resulted in no detectable phenotypes, but sterility was observed for *hsp-1* and *vps-35*. However, the frameshift mutations observed at *hsp-1* and *vps-35* in the MA lines might result in the expression of truncated protein products and result in less severe phenotypes than what is observed in RNAi studies. The largest indel observed was a ~500-bp insertion in MA46 at locus Y41D4B (Supplementary Fig. S2). Although exhaustive efforts to determine the sequence of this mutation were unsuccessful, the presence of five copies of the *C. elegans* repetitive element Rc123 (ref. 21) at this locus in wild-type (N2) DNA indicates that there might have been an expansion of these elements at this locus in MA46. Size variation at this locus between the *C. elegans* natural isolates is due to variation in Rc123 copy number.

Base substitutions are about fourfold more frequent than indels in the *C. elegans* natural isolates⁷, and this observation is often used to justify the use of base-substitution-inducing mutagens such as ethylmethane sulphonate to mimic spontaneous mutation processes^{22,23}. However, our observation that the mutation rate to indels is higher than the rate for base substitutions in the MA lines raises questions about the validity of this assumption. It is unlikely that any single mutagen would induce the diverse assortment of spontaneous mutations observed in the MA lines.

Transitions are consistently more common than transversions in natural patterns of nuclear variation in metazoans, but it remains unclear whether this pattern is driven by biased mutation and/or selection. Among the 13 base substitutions observed in the MA lines (Table 2), 8 were transitions and 5 were transversions, yielding a transition/transversion (Ts/Tv) ratio (1.6) that is strikingly similar to that observed at the same nuclear loci in the *C. elegans* natural isolates (1.7) (ref. 7). Even higher transition biases were found in the mitochondrial genomes of both the *C. elegans* MA lines (Ts/Tv = 4.3)⁸ and natural isolates (Ts/Tv = 2.6)⁷. These observations suggest that underlying mutational forces drive the widely observed transition-bias phenomenon in both the nuclear and mitochondrial genomes of *C. elegans*.

Last, we note that our sequence-based estimate of U_t (~2.1 mutations per haploid genome per generation) is about two orders

of magnitude higher than the haploid deleterious genomic mutation rate (U_d) estimated by laboratory fitness assays with the same set of *C. elegans* MA lines ($U_d \approx 0.015$ mutations per haploid genome per generation)⁶, indicating that most mutations in the MA lines (more than 99%) might either be neutral or have deleterious effects too mild to be detected in laboratory fitness assays. However, comparative analyses of *Caenorhabditis* genome sequences suggest that the vast majority of nonsynonymous mutations (~94% on the basis of a between-species K_a/K_s ratio of 0.06)²⁴ are sufficiently deleterious for selection to act against them in nature, and fitness-based estimates of U_d are known to be highly dependent on environmental conditions²⁵. If we consider only nonsynonymous base substitution mutations and indels in exon sequence, the mutation rate reported here coupled with the K_a/K_s data for *Caenorhabditis* yields an estimate of 0.48 for U_d (see Methods). This approximation of U_d is probably an underestimate because it does not consider mutations in intron and intergenic regions that can also have negative effects. Nevertheless, the fact that our sequence-based minimal estimate for U_d in *C. elegans* is about 30-fold higher than the estimate based on laboratory fitness assays indicates that most spontaneous mutations in *C. elegans* have very mildly deleterious effects that are realized only in natural contexts. □

Methods

Mutation detection and confirmation

Details about PCR amplification and sequencing of *C. elegans* nuclear loci have been described previously^{7,8}. Most loci sequenced were randomly distributed across *C. elegans* chromosomes by designing PCR primer pairs around chromosomal positions selected by a random-number generator. All amplifications were performed with a large amount of genomic DNA (~25,000 diploid genomes per reaction) and 2 U Taq DNA polymerase (Applied Biosystems) to eliminate artefacts associated with initial amplification from small amounts of genomic DNA. All PCR products were directly sequenced in both directions, and internal primers were used where necessary.

DNA sequence text files were aligned to wild-type (N2) sequences with CLUSTALW²⁶ to identify putative mutations in the MA lines. Putative mutations identified in the alignments were then scrutinized visually on the electropherogram data to eliminate base-caller errors and other sequencing artefacts. All putative mutations showing evidence of wild-type or ambiguous sequence data were resequenced. Putative mutations supported by clean, unambiguous electropherogram data were then evaluated on the opposite strand (sequencing reaction in the opposite direction), with internal primers where necessary. If we conservatively assume that our sequencing error rate was on the high end of that reported for the human genome (~10⁻⁴) (ref. 27) then an error rate of ~10⁻⁸ is expected for our approach where mutations were confirmed on both strands. Assuming this conservative error rate, we would expect only ~0.04 mutations due to sequencing error in the more than 4 megabases of DNA that we surveyed. All mutations reported here received high Phred scores (more than 33) on both strands (corresponding to an error probability of less than 0.05%)²⁸. Finally, starting from genomic DNA we re-amplified and sequenced a subset (10 of 30) of the PCR products containing the initially identified mutations, and in each of the ten instances the mutations were again detected.

Calculation of mutation rates

Individual mutation rates were calculated with the equation $\mu = m/(LnT)$, where μ is the mutation rate (per nucleotide site per generation), m is the number of observed mutations, L is the number of MA lines, n is the number of nucleotide sites and T is the time in generations, as described previously⁸. The standard errors for individual mutation rates were calculated as $[\mu/(LnT)]^{1/2}$, as described previously⁸. Three highly similar independent mutation rates were calculated for each of the three mutation assays: 2.0×10^{-8} mutations per site per generation ($\pm 5.8 \times 10^{-9}$) for the first assay (280 generations), 2.6×10^{-8} ($\pm 8.6 \times 10^{-9}$) for the second assay (353 generations) and 2.1×10^{-8} ($\pm 7.0 \times 10^{-9}$) for the third assay (396 generations). The single rate reported in the text is the average of the three rates, weighted by the inverse of the sampling variance for each rate. To avoid providing an upwardly biased estimate of the mutation rate, we excluded mutations detected at homopolymeric nucleotide runs of 8 bp or longer that are hotspots for indel mutation and were specifically targeted for another study²⁹.

Poisson distribution of mutations

For both the distribution of mutations between individual MA lines and between individual PCR products, our observations were not significantly different from Poisson expectations. Assuming a Poisson distribution and a mean of 0.45 mutations per MA line (with a weighted average of 66 total MA lines), the expected number of MA lines with 0, 1, 2 and 3 mutations are 42, 19, 4 and 1, respectively; we observed 43, 17, 5 and 1 MA lines. Assuming a Poisson distribution and a mean of 0.39 mutations per PCR product locus surveyed (76 total), the expected number of loci with 0, 1, 2 and 3 mutations are 51, 20, 4 and 1, respectively; we observed 54, 15, 5 and 2 loci.

Estimation of the deleterious genomic mutation rate

If we use $1-(K_a/K_s)$ as a measure of the fraction of mutations in protein-coding genes that are deleterious (for *Caenorhabditis*, $K_a/K_s \approx 0.06$)²⁴ and multiply this by the genomic rate of nonsynonymous base substitution mutations in exon sequences in the MA lines (0.18 nonsynonymous mutations per genome per generation, given that ~26% of the *C. elegans* genome is exon, ~75% of exon nucleotides are nonsynonymous sites, and our total mutation rate of 0.9 base substitutions per genome per generation), we get 0.17 as an estimate of U_d for *C. elegans*. Adding the rate of indel mutations in exon sequence (0.31 indel mutations in exons per genome per generation, given that ~26% of the *C. elegans* genome is exon and our total mutation rate of 1.2 indels per genome per generation) and assuming all exon indels are deleterious, yields 0.48 as an estimate of U_d for *C. elegans*.

Received 3 February; accepted 1 June 2004; doi:10.1038/nature02697.

- Drake, J. W., Charlesworth, B., Charlesworth, D. & Crow, J. F. Rates of spontaneous mutation. *Genetics* **148**, 1667–1686 (1998).
- Nachman, M. W. & Crowell, S. L. Estimate of the mutation rate per nucleotide in humans. *Genetics* **156**, 297–304 (2000).
- Kumar, S. & Subramanian, S. Mutation rates in mammalian genomes. *Proc. Natl Acad. Sci. USA* **99**, 803–808 (2002).
- Robertson, H. M. The large *srh* family of chemoreceptor genes in *Caenorhabditis* nematodes reveals processes of genome evolution involving large duplications and deletions and intron gains and losses. *Genome Res.* **10**, 192–203 (2000).
- Witherspoon, D. J. & Robertson, H. M. Neutral evolution of ten types of mariner transposons in the genomes of *Caenorhabditis elegans* and *Caenorhabditis briggsae*. *J. Mol. Evol.* **56**, 751–769 (2003).
- Vassilieva, L. L., Hook, A. M. & Lynch, M. The fitness effects of spontaneous mutations in *Caenorhabditis elegans*. *Evolution* **54**, 1234–1246 (2000).
- Denver, D. R., Morris, K. & Thomas, W. K. Phylogenetics in *Caenorhabditis elegans*: an analysis of divergence and outcrossing. *Mol. Biol. Evol.* **20**, 393–400 (2003).
- Denver, D. R., Morris, K., Lynch, M., Vassilieva, L. L. & Thomas, W. K. High direct estimate of the mutation rate in the mitochondrial genome of *Caenorhabditis elegans*. *Science* **289**, 2342–2344 (2000).
- Ochman, H. Neutral mutations and neutral substitutions in bacterial genomes. *Mol. Biol. Evol.* **20**, 2091–2096 (2003).
- Petrov, D. A., Lozovskaya, E. R. & Hartl, D. L. High intrinsic rate of DNA loss in *Drosophila*. *Nature* **384**, 346–349 (1996).
- Petrov, D. A. & Hartl, D. L. Pseudogene evolution and natural selection for a compact genome. *J. Hered.* **91**, 221–227 (2000).
- Hirotsune, S. *et al.* An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene. *Nature* **423**, 91–96 (2003).
- Yamada, K. *et al.* Empirical analysis of transcriptional activity in the *Arabidopsis* genome. *Science* **302**, 842–846 (2003).
- Balakirev, E. S. & Ayala, F. J. Pseudogenes: are they 'junk' or functional DNA? *Annu. Rev. Genet.* **37**, 123–151 (2003).
- Charlesworth, B. The changing sizes of genes. *Nature* **384**, 315–316 (1996).
- Lynch, M. & Conery, J. S. The origins of genome complexity. *Science* **302**, 1401–1404 (2003).
- Marais, G., Mouchiroud, D. & Duret, L. Does recombination improve selection on codon usage? Lessons from nematode and fly complete genomes. *Proc. Natl Acad. Sci. USA* **98**, 5688–5692 (2001).
- Hahn, M. W., Stajich, J. E. & Wray, G. A. The effects of selection against spurious transcription factor binding sites. *Mol. Biol. Evol.* **20**, 901–906 (2003).
- Langley, C. H. & Ito, K. Spontaneous mutability in *Drosophila melanogaster*, in natural and laboratory environments. *Mutat. Res.* **36**, 385–386 (1976).
- Gunsalus, K. C., Yueh, W. C., MacMenamin, P. & Piano, F. RNAiDB and PhenoBlast: web tools for genome-wide phenotypic mapping projects. *Nucleic Acids Res.* **32**, D406–D410 (2004).
- Naclerio, G. *et al.* Molecular and genomic organization of clusters of repetitive DNA sequences in *Caenorhabditis elegans*. *J. Mol. Biol.* **226**, 159–168 (1992).
- Keightley, P. D. & Ohnishi, O. EMS-induced polygenic mutation rates for nine quantitative characters in *Drosophila melanogaster*. *Genetics* **148**, 753–766 (1998).
- Davies, E. K., Peters, A. D. & Keightley, P. D. High frequency of cryptic deleterious mutations in *Caenorhabditis elegans*. *Science* **285**, 1748–1751 (1999).
- Stein, L. D. *et al.* The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics. *PLoS Biol.* **1**, 166–192 (2003).
- Kondrashov, A. S. & Houle, D. Genotype–environment interactions and the estimation of the genomic mutation rate in *Drosophila melanogaster*. *Proc. R. Soc. Lond. B* **258**, 221–227 (1994).
- Higgins, D. G., Thompson, J. D. & Gibson, T. J. Using CLUSTAL for multiple sequence alignments. *Methods Enzymol.* **266**, 383–402 (1994).
- Hill, F., Gemund, C., Benes, V., Ansorge, W. & Gibson, T. J. An estimate of large-scale sequencing accuracy. *EMBO Rep.* **1**, 29–31 (2000).
- Richterich, P. Estimation of errors in 'raw' DNA sequences: a validation study. *Genome Res.* **8**, 251–259 (1998).
- Denver, D. R. *et al.* Abundance, distribution, and mutation rates of homopolymeric nucleotide runs in the genome of *Caenorhabditis elegans*. *J. Mol. Evol.* **58**, 584–595 (2004).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank L. L. Vassilieva, S. Estes, V. Katju and C. Steding for their respective roles in propagating and maintaining the MA lines over the past 5 years; D. Ash for help with primer sequence design and DNA sequencing; and the *Caenorhabditis* Genetics Center for providing the *C. elegans* natural isolates. This work was supported by a University of Missouri Research Board grant to W.K.T., and an NIH grant to M.L. and W.K.T.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.R.D. (ddenver@bio.indiana.edu).

Optimal neural population coding of an auditory spatial cue

Nicol S. Harper^{1,2} & David McAlpine¹

¹Department of Physiology and UCL Ear Institute and ²CoMPLEX, University College London, London WC1E 6BT, UK

A sound, depending on the position of its source, can take more time to reach one ear than the other. This interaural (between the ears) time difference (ITD) provides a major cue for determining the source location^{1,2}. Many auditory neurons are sensitive to ITDs^{3,4}, but the means by which such neurons represent ITD is a contentious issue. Recent studies question whether the classical general model (the Jeffress model⁵) applies across species^{6,7}. Here we show that ITD coding strategies of different species can be explained by a unifying principle: that the ITDs an animal naturally encounters should be coded with maximal accuracy. Using statistical techniques and a stochastic neural model, we demonstrate that the optimal coding strategy for ITD depends critically on head size and sound frequency. For small head sizes and/or low-frequency sounds, the optimal coding strategy tends towards two distinct sub-populations tuned to ITDs outside the range created by the head. This is consistent with recent observations in small mammals^{6,7}. For large head sizes and/or high frequencies, the optimal strategy is a homogeneous distribution of ITD tunings within the range created by the head. This is consistent with observations in the barn owl^{8–10}. For humans, the optimal strategy to code ITDs from an acoustically measured distribution depends on frequency; above 400 Hz a homogeneous distribution is optimal, and below 400 Hz distinct sub-populations are optimal.

The ability to localize sound sources has obvious survival value, whether for prey or predator. Many vertebrates, including humans, make use of ITDs to localize sounds in the horizontal plane. These ITDs can be in the order of just a few tens of microseconds. In the Jeffress model⁵, these minute ITDs are encoded by an array of coincidence-detector neurons. Each neuron is tuned for (responds maximally to) an ITD within the range created by the head size (the physiological range), the precise tuning being determined by the difference in axonal conduction time from each ear. The Jeffress model, developed with human spatial hearing in mind, has received extensive support from studies in barn owls^{8–10}, and until recently was presumed to apply to mammals¹¹. However, the preferred ITD tuning of neurons recorded in small mammals that use ITDs for sound localization seems to lie outside the physiological range^{6,7}. This is markedly different to the preferred ITD tuning observed in the barn owl and suggested by the Jeffress model. It is now debated whether a single unifying principle or model can any longer explain the means by which ITD is encoded across the wide range of species in which ITD sensitivity is observed.

We investigated the possibility that different coding strategies observed in different species can be explained by the demand for accurate ITD coding. For simplicity, we considered first how to encode most accurately the ITDs in the ongoing fine structure of a pure tone using a population of ITD-tuned model neurons. In mammals, fine-structure ITD sensitivity is restricted to sound frequencies below ~1,500 Hz. At higher frequencies, some ITD information can be conveyed by the envelope structure of complex sounds¹²; however, this seems to have relatively little impact on localization judgements^{2,13}, and thus we only consider fine-structure ITDs.

Here, we present the optimal coding strategies for four species with different head sizes and/or different sound-frequency ranges

over which ITD sensitivity is observed: gerbil, a species with low-frequency ($< \sim 1,500$ Hz) ITD sensitivity and a small head; barn owl, a high-frequency ($< \sim 10$ kHz) ITD specialist with a small head; human, low-frequency ITD sensitivity with a large head; and cat, low-frequency ITD sensitivity with an intermediate head size.

We use the simplest neural model, the ITD tuning curve (that is, a function describing a neuron's mean firing rate in response to different ITDs). For pure tones, each ITD corresponds to an ongoing interaural phase difference (IPD). IPDs up to half the pure-tone period in magnitude constitute ITD as a proportion of that period. Because pure tones are periodic, IPDs beyond half a cycle of the period are subject to 'phase wrapping'. For example, an IPD of $+0.6$ cycles is equivalent to an IPD of -0.4 cycles. Although measured ITD tuning curves scale with frequency, when considered in terms of IPD they are largely invariant with frequency, being bell-shaped around a 'best IPD' (that is, the IPD that evokes maximum discharge rate; Fig. 1a). Thus, for convenience, all ITD tuning curves are represented in terms of the equivalent IPD. We approximate the IPD tuning curves with a cosine-based function (see Methods).

We assume that IPD is encoded in the neural population's spike counts within a time window, with intrinsic Poisson noise producing variability in the spike counts. The overall measure of the population's coding error is the sum of the error for each IPD, weighted by the probability of occurrence of that IPD. The measure of coding error for each IPD is the variance of the population's representation of that IPD, as calculated from Fisher information (see Methods). For each species, for each pure-tone frequency, we determine the optimal coding strategy (optimal distribution of best IPDs) by shifting systematically the best IPD of each neuron (in a population of 200) until the overall coding error is minimized.

For the probability distribution of the IPDs of a pure tone we observe that the maximum ITD an animal experiences under natural listening conditions is limited by the size of its head. The maximum IPD corresponds to the maximum ITD as a proportion of the pure-tone period, up to a limit of half a cycle. We assume that only IPDs with magnitude smaller than the maximum IPD (that is, within the physiological range) are encountered, and that all IPDs within this range are equally likely. The effect of phase wrapping on the IPD probability distribution is ignored, as including it in the model has little qualitative effect on the results. The physiological range for each tone frequency is illustrated in Fig. 1b for the cat.

The first species considered, the gerbil, has a relatively small head with a maximum ITD of approximately $\pm 120 \mu\text{s}$. Consequently, for every frequency at which IPD sensitivity is observed, the gerbil's physiological range of IPDs is narrow relative to the IPD tuning curve. The optimal coding strategy for a 500-Hz pure tone (Fig. 2a) takes the form of two distinct neural sub-populations, with best IPDs positioned either side of zero and beyond the gerbil's physiological range. These two sub-populations are apparent over the full range of ITD-sensitive tone frequencies (Fig. 2c), with only a slight divergence at the highest frequencies. This is consistent with recent

electrophysiological recordings made in the brainstem of the gerbil⁷ and the midbrain of the guinea-pig^{6,14,15}, another small mammal with low-frequency IPD sensitivity. It is clear that this solution is optimal from the Fisher information, a local measure of coding accuracy, which is highest over the slopes of the IPD tuning curves rather than at their peaks. To position the peak Fisher information within the physiological range it is therefore necessary to position peaks of IPD functions beyond the physiological range.

The solution for the gerbil could be taken to suggest that the optimal coding strategy is largely independent of frequency and involves distinct neural sub-populations. However, the solution for the barn owl indicates that this is not the case. Barn owls, with a similar maximum ITD ($\pm 180 \mu\text{s}$) to the gerbil, make use of fine-structure ITD cues over a much higher frequency range ($3\text{--}10$ kHz^{8–10}) than mammals. Over this range, the physiological range of IPDs (Fig. 2b) is wide relative to the IPD tuning curves, and the optimal coding strategy involves, instead, a homogeneous distribution of neurons tuned to IPDs within the physiological range.

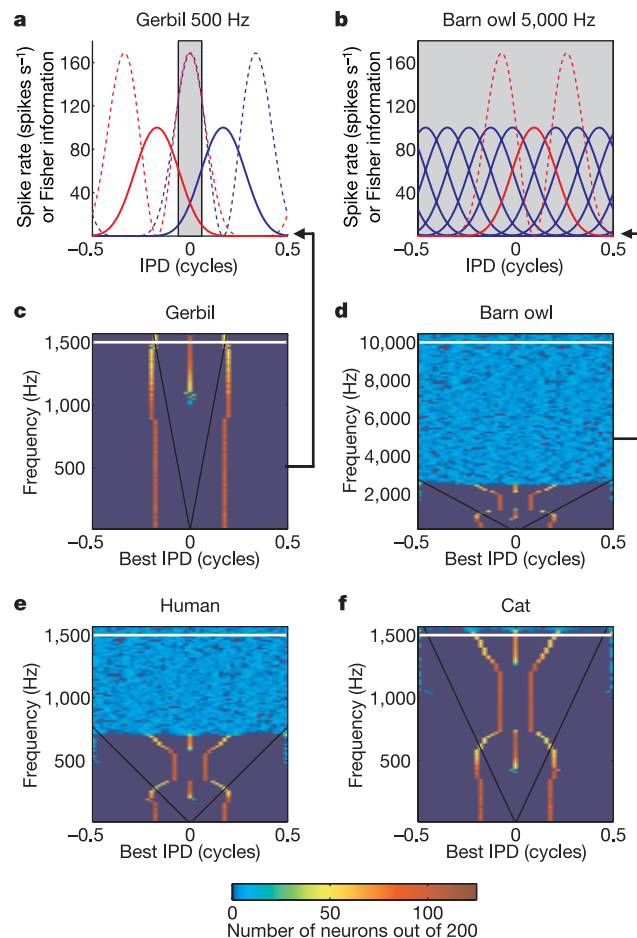


Figure 2 Optimal distributions of tuning curves for coding IPDs. **a**, Optimal distribution of tuning curves to code the IPD of 500-Hz tones in gerbil for left (red) and right (blue) sub-populations. Dashed lines show Fisher information curves (scaled by dividing by duration of the spike-count window, T). The grey region indicates the physiological range of IPDs, whereas vertical black lines indicate the maximum IPD. **b**, Optimal distribution for coding IPDs of 5,000-Hz tones in barn owl (sample only shown). Red lines show tuning curve (solid) and scaled Fisher information (dashed) for one neuron. **c–f**, Optimal distributions for coding pure tone IPDs in gerbil (**c**), barn owl (**d**), human (**e**) and cat (**f**). The abscissa and ordinate indicate best IPD and pure-tone frequency, respectively. For each frequency, the colour indicates the number of neurons (of 200) with that best IPD. White lines indicate frequency limit for IPD sensitivity, and black lines maximum IPD. Black arrows show how **a** and **b** illustrate the distribution at particular frequency bands in **c** and **d**, respectively.

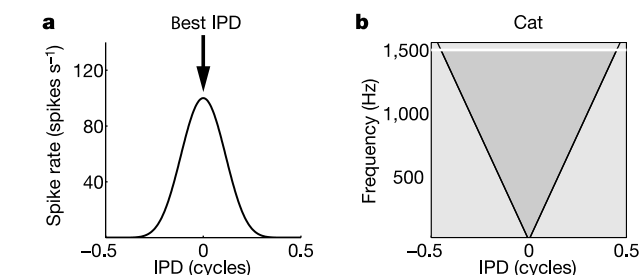


Figure 1 Key features of model. **a**, Illustration of the IPD tuning curve used in the model. **b**, Physiological range (dark grey) of IPD in a cat, showing the limit of IPD sensitivity (white line) and the maximum IPD for each frequency (black line).

This is consistent with physiological recordings made at all levels of the barn owl binaural pathway^{8–10}. When the effects of phase wrapping on the IPD distribution are included in the analysis, the homogeneous distribution is maintained, save for slight fluctuations in neuron densities. The homogeneous distribution seems to be consistent with the basic tenets of the classical Jeffress model⁵. The solution is optimal because the slopes of IPD tuning curves, where Fisher information is maximal, always reside within the physiological range, and homogeneity is preferred because of the reciprocal term in equation (2) (see Methods). Previous general theoretical studies using relatively narrow tuning curves also

demonstrate that the optimal distribution of preferred tuning follows the probability distribution of the stimulus¹⁶.

The homogeneous solution departs from the Jeffress model in the form of the presumed coding strategy. The Jeffress model posits local coding for ITD, whereas we assume a population code, in which the slopes of functions are critical. This is consistent with a recent study of the barn owl midbrain¹⁷, which reports neural resolution of a parameter related to ITD (azimuthal location of broadband noise) to be highest over the slopes, rather than the peaks, of functions. Furthermore, although the homogeneous distribution is the optimal strategy above 3 kHz in the barn owl, below 3 kHz the solution is more complex, and distinct sub-populations are observed (Fig. 2d). Below approximately 800 Hz, the solution is essentially the same as for the gerbil. There is a paucity of data for barn owl neurons with sound frequency tuning below 3 kHz. However a recent study¹⁸ reports a population of ITD-sensitive neurons with tuning for sound frequency extending down to the sub-kHz range. Although this study indicates a sub-population of neurons tuned around zero ITD, a distinct sub-population also exists with response maxima to ITDs significantly beyond the barn owl's physiological range. Distinct sub-populations and tuning beyond the physiological range are inconsistent with the Jeffress model; however, they are consistent with the optimal coding model.

The optimal coding strategies described thus far consist largely of either two distinct, opposed sub-populations tuned to ITDs beyond the physiological range, or a single homogeneous distribution lying entirely within the physiological range. For humans, however, with their relatively large head size and maximum ITD in excess of $\pm 600 \mu\text{s}$, the solution is more variable (Fig. 2e). For the lowest frequencies (<250 Hz) the optimal coding strategy is the same as for the gerbil, a distribution consisting of two distinct sub-populations. For the highest frequencies (>700 Hz) the homogeneous distribution is optimal. For intermediate frequencies (250–700 Hz), however, a transition occurs between the two coding strategies, with the optimal strategy consisting of several distinct sub-populations. Similarly, the solution for the cat (Fig. 2f), with a physiological range of ITDs intermediate to that of gerbil and humans, emphasizes the transition distributions.

The model indicates that a critical factor in determining the strategy for optimal coding is the relative width of the physiological range of IPDs compared with the width of individual IPD tuning curves. The model assumes the shape of the IPD tuning curve to be similar across species. Therefore, the exact form of the optimal coding strategy depends only on the maximum IPD at that frequency. As maximum IPD scales with frequency and/or head size, the panels of Fig. 2c–f are simply scaled and shifted versions of each other. However, it is likely that the optimal strategy will also depend on the exact shape of the IPD probability distribution, and to this end we measured the probability distribution of IPDs for human listeners and applied this distribution to the model. Small microphones, positioned at the entrance to each ear canal, were used to make stereo sound recordings from three human subjects interacting in a range of acoustic environments. IPD probability distributions were calculated for each frequency component in recordings from each subject–environment combination (Figs 3a–d, see Methods). The pure-tone model can be used to find the optimal coding strategy for these IPD distributions if it is assumed that ITD coding in each frequency band is optimized independently, a reasonable assumption as auditory processing is carried out within finite frequency channels up to a very high level in the auditory pathway. Figure 3e–h illustrates the corresponding solutions for each subject–environment combination. Above 400 Hz, the optimal coding strategy is a homogeneous distribution, consistent with the previous assumption of equal probability IPDs. Despite the IPD distributions being subject to the effects of phase wrapping, only small fluctuations in the homogeneous distribution were observed.

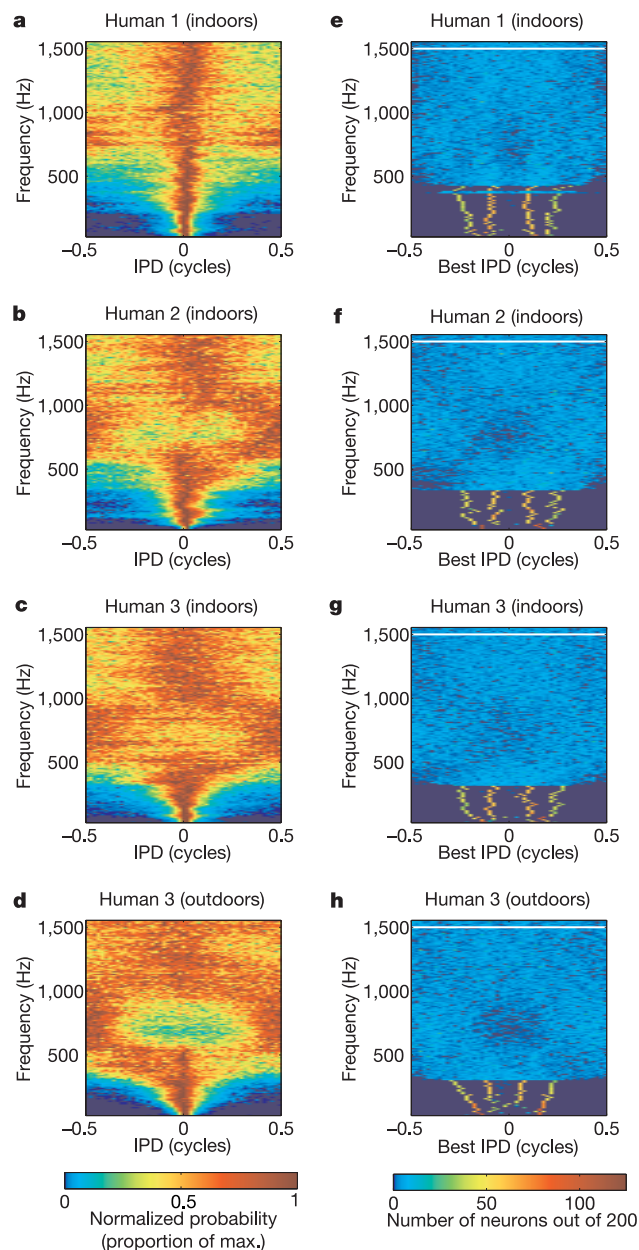


Figure 3 Probability distributions of IPDs at different frequencies for three human subjects. **a–d**, Subjects in a room (**a–c**), and outdoors (**d**; same subject as in **c**). The IPD distribution at each frequency is normalized to maximum. The abscissa and ordinate indicate IPD and frequency, respectively. Colour indicates the normalized probability. **e–h**, Optimal distributions of best IPDs for **a–d**. The abscissa and ordinate indicate best IPD and frequency, respectively. For each frequency, the colour indicates the number of neurons (out of 200) with that best IPD. White lines indicate the frequency limit of IPD sensitivity.

Below 400 Hz, however, a modification of the two sub-population solution is optimal, with two central sub-populations and two additional, less populous, flanking sub-populations.

Although we assume ITD coding to be optimized independently for each frequency band, similar solutions are found without this assumption. If we optimize a population of model neurons to encode most accurately the ITD of a broadband noise, with the ITD tuning of each neuron modelled by a Gabor function with randomly assigned frequency, the solution is consistent with the pure-tone model (data not shown). The two sub-population solution was obtained for low-frequency-tuned neurons and small head sizes. For larger head sizes and higher frequency tuned neurons, the solution tended towards the homogeneous distribution. Furthermore, if the measure of coding accuracy is based on mutual information¹⁶ rather than expected variance, it has little qualitative effect on the solutions of the models. Finally, as long as spike-rate variance is approximately proportional to mean spike rate, the same coding strategies result.

A question arises as to the developmental mechanism by which each optimal distribution could be achieved. Where the model suggests only one coding strategy across the entire frequency range at which ITDs are used (for example, 3–10 kHz in the barn owl) the optimal distribution could be hardwired, possibly by means of neural delay lines¹⁹. Another, more dynamic, mechanism is suggested by recent evidence from the gerbil of a fast glycine-mediated input that modulates the best ITD of medial superior olive neurons⁷. Evidence suggests that this inhibitory input can be modulated over development by the sound environment²⁰. Such a regulatory mechanism for ITD tuning would be useful where developmental changes in head size influence which coding strategy is optimal, or where, as we predict for humans, different coding strategies are optimal for different sound frequencies.

The relatively simple model we describe provides an explanation for reported differences in neural coding of ITDs between different mammals, and between mammals and barn owls^{6–10,18}. Assuming only the requirement of accurate coding of this auditory spatial cue, the model suggests a general framework in which to predict the coding strategy adopted by any species, depending on head size and ITD-sensitive sound frequency range. Two general categories of coding strategy were apparent: one with distinct sub-populations of ITD-sensitive neurons, and one with a homogeneous population. The model also predicts that some species, humans for example, may use different ITD-coding strategies over different sound frequency ranges. The assumption of accurate coding provides a unifying principle by which future investigations into the anatomical and physiological mechanisms responsible for ITD sensitivity may be guided. □

Methods

The error measure

The error measure $V(\varphi)$ is minimized by varying the best IPDs, φ , of a population of model IPD-sensitive neurons, to provide for the most accurate distribution of best IPDs, φ_{optimal} . $V(\varphi)$ is the expected value of the variance $\sigma^2(\theta|\varphi)$ of the representation of IPD θ by the spike counts of the model neurons, over a sufficiently long time window, given IPD probability distribution $p(\theta)$. The variance is small enough (just-noticeable difference in IPD typically $\sim 3^\circ$)²¹ to ignore the angular nature of IPD θ (angles θ and φ here are in radians). Thus:

$$V(\varphi) = \int_{-\pi}^{\pi} \sigma^2(\theta|\varphi) p(\theta) d\theta \quad (1)$$

For each IPD, a lower bound on the variance $\sigma^2(\theta|\varphi)$ in the coded IPD is given by the Cramér–Rao bound $\sigma^2(\theta|\varphi) \geq 1/F(\theta|\varphi)$, where $F(\theta|\varphi)$ is the Fisher information of the population²². In the limit of a large number of N neurons the Cramér–Rao bound becomes an equality, and any bias tends to zero. Hence:

$$V(\varphi) = \int_{-\pi}^{\pi} F(\theta|\varphi)^{-1} p(\theta) d\theta \quad (2)$$

Assuming that the noise between neurons is not correlated, the Fisher information $F(\theta|\varphi)$ of the population is the sum of the Fisher information $f(\theta|\varphi_i)$ of each individual neuron i , where φ_i is a neuron's best IPD. $f(\theta|\varphi_i)$, for intrinsic Poisson noise, in turn depends on the IPD tuning curve, $g(\theta|\varphi_i)$, and the duration T of the spike-count time window. Thus the

population Fisher information is given by:

$$F(\theta|\varphi) = \sum_{i=1}^N f(\theta|\varphi_i), \text{ with } f(\theta|\varphi_i) = T \frac{[\frac{\partial}{\partial \theta} g(\theta|\varphi_i)]^2}{g(\theta|\varphi_i)} \quad (3)$$

$V(\varphi)$ is minimized using a conjugate-gradient-based algorithm (Minimize; <http://www.kyb.tuebingen.mpg.de/bs/people/carl/code/>). It finds the minimum within 1,000 line searches, and usually fewer than 50. φ is initially normally distributed around 0 IPD, however the same solutions result when other initial distributions are used.

IPD tuning curves

For the IPD tuning curve we chose the function $g(\theta|\varphi_i)$ to provide a reasonable fit to the inner slope of a number of tuning curves recorded from the guinea-pig inferior colliculus. The exact function is:

$$g(\theta|\varphi_i) = r_{\text{max}} \left[\frac{1}{2} + \frac{1}{2} \cos(\theta - \varphi_i) \right]^4 \quad (4)$$

The maximum firing rate r_{max} , like the window length T , is a constant that can be ignored in the optimization so long as $r_{\text{max}}T$ is sufficiently large²³. The maximum firing rates tend to be high in the auditory brainstem (>100 spikes per second), and psychophysically derived binaural integration times are long (~ 100 ms)²⁴.

Measuring the probability distribution of IPDs

Small microphones (Knowles FG3452) were fixed at the entrance to each ear canal in three human subjects and ambient sounds recorded over approximately 2 min. The subject–environment combinations were male 1 (Fig. 3a) or female 1 (Fig. 3b) in conversation with the experimenter walking around a room (4 × 4 × 3 m), male 2 walking in the same environment, but in the absence of conversation (Fig. 3c), and male 2 seated in the front quadrangle (100 × 50 m) of University College London, with multiple other persons in the quadrangle (Fig. 3d). For each subject–environment combination the recorded stereo sound was partitioned into sections using a 100-ms (seven standard deviations) gaussian window, shifted in 50-ms increments. A Fast Fourier transform was applied to each section for each ear, to give the phase for each frequency band. From this, the IPD for each frequency band in each section was calculated as the difference in phase between the ears. The probability distribution of IPDs across all of the sections was then obtained for each frequency band.

Received 25 February; accepted 18 June 2004; doi:10.1038/nature02768.

- Rayleigh, L. On our perception of sound direction. *Phil. Mag.* **13**, 214–232 (1907).
- Wightman, F. L. & Kistler, D. J. The dominant role of low-frequency interaural time differences in sound localisation. *J. Acoust. Soc. Am.* **91**, 1648–1661 (1992).
- Kuwada, S. & Yin, T. C. T. Binaural interaction in low-frequency neurons in inferior colliculus of the cat. I. Effects of long interaural delays, intensity, and repetition rate on interaural delay function. *J. Neurophysiol.* **50**, 981–999 (1983).
- Yin, T. C. T. & Chan, J. C. K. Interaural time sensitivity in medial superior olive of cat. *J. Neurophysiol.* **64**, 465–488 (1990).
- Jeffress, L. A. A place theory of sound localisation. *J. Comp. Physiol. Psychol.* **41**, 35–39 (1948).
- McAlpine, D., Jiang, D. & Palmer, A. R. A neural code for low-frequency sound localization in mammals. *Nature Neurosci.* **4**, 396–401 (2001).
- Brand, A., Behrend, O., Marquardt, T., McAlpine, D. & Grothe, B. Precise inhibition is essential for microsecond interaural time difference coding. *Nature* **417**, 543–547 (2002).
- Takahashi, T. & Konishi, M. Selectivity for interaural time difference in the owl's midbrain. *J. Neurosci.* **6**, 3413–3422 (1986).
- Wagner, H., Takahashi, T. & Konishi, M. Representation of interaural time difference in the central nucleus of the barn owl's inferior colliculus. *J. Neurosci.* **7**, 3105–3116 (1987).
- Coles, R. B. & Guppy, A. Directional hearing in the barn owl (*Tyto alba*). *J. Comp. Physiol. A* **163**, 117–133 (1988).
- Joris, P. X., Smith, P. H. & Yin, T. C. T. Coincidence detection in the auditory system: 50 years after Jeffress. *Neuron* **21**, 1235–1238 (1998).
- Henning, G. B. Detectability of interaural delay in high-frequency complex waveforms. *J. Acoust. Soc. Am.* **55**, 84–90 (1974).
- Bernstein, L. R. & Trahiotis, C. Lateralization of sinusoidally amplitude-modulated tones: Effects of spectral locus and temporal variation. *J. Acoust. Soc. Am.* **78**, 514–523 (1985).
- Skottun, B. C., Shackleton, T. M., Arnott, R. H. & Palmer, A. R. The ability of inferior colliculus neurons to signal differences in interaural delay. *Proc. Natl Acad. Sci. USA* **98**, 14050–14054 (2001).
- Shackleton, T. M., Skottun, B. C., Arnott, R. H. & Palmer, A. R. Interaural time difference discrimination thresholds for single neurons in the inferior colliculus of Guinea pigs. *J. Neurosci.* **23**, 716–724 (2003).
- Brunel, N. & Nadal, J. P. Mutual information, Fisher information, and population coding. *Neural Comput.* **10**, 1731–1757 (1998).
- Bala, A. D., Spitzer, M. W. & Takahashi, T. T. Prediction of auditory spatial acuity from neural images on the owl's auditory space map. *Nature* **424**, 771–774 (2003).
- Wagner, H., Mazer, J. A. & von Campenhausen, M. Response properties of neurons in the core of the central nucleus of the inferior colliculus of the barn owl. *Eur. J. Neurosci.* **15**, 1343–1352 (2002).
- Carr, C. E. & Konishi, M. Axonal delay lines for time measurement in the owl's brainstem. *Proc. Natl Acad. Sci. USA* **85**, 8311–8315 (1988).
- Kapfer, C., Seidl, A. H., Schweizer, H. & Grothe, B. Experience-dependent refinement of inhibitory inputs to auditory coincidence-detector neurons. *Nature Neurosci.* **5**, 247–253 (2002).
- Domnitz, R. H. & Colburn, H. S. Lateral position and interaural discrimination. *J. Acoust. Soc. Am.* **61**, 1586–1598 (1977).
- Dayan, P. & Abbot, L. F. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems* (MIT Press, Cambridge, Massachusetts, 2001).
- Bethge, M., Rotermund, D. & Pawelzik, K. Optimal short-term population coding: When Fisher information fails. *Neural Comput.* **14**, 2317–2351 (2002).

24. Culling, J. F. & Summerfield, Q. Measurements of the binaural temporal window using a detection task. *J. Acoust. Soc. Am.* **103**, 3540–3553 (1998).

Acknowledgements We thank C. Rasmussen for the algorithm 'Minimize'. This work was supported by a MRC Career Establishment Grant to D.M. and a MRC Bioinformatics Studentship to N.S.H.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.M. (d.mcalpine@ucl.ac.uk).

Sirtuin activators mimic caloric restriction and delay ageing in metazoans

Jason G. Wood¹*, Blanka Rogina²*, Siva Lavu¹, Konrad Howitz³, Stephen L. Helfand², Marc Tatar⁴ & David Sinclair¹

¹Department of Pathology, Harvard Medical School, 77 Ave. Louis Pasteur, Boston, Massachusetts 02115, USA

²Department of Genetics and Developmental Biology, University of Connecticut Health Center, 263 Farmington Ave., Farmington, Connecticut 06030, USA

³BIOMOL Research Laboratories, Inc., 5120 Butler Pike, Plymouth Meeting, Pennsylvania 19462, USA

⁴Department of Ecology and Evolutionary Biology, Box G-W, Brown University, Providence, Rhode Island 02912, USA

* These authors contributed equally to this work

Caloric restriction extends lifespan in numerous species. In the budding yeast *Saccharomyces cerevisiae* this effect requires Sir2 (ref. 1), a member of the sirtuin family of NAD⁺-dependent deacetylases^{2,3}. Sirtuin activating compounds (STACs) can promote the survival of human cells and extend the replicative lifespan of yeast⁴. Here we show that resveratrol and other STACs activate sirtuins from *Caenorhabditis elegans* and *Drosophila melanogaster*, and extend the lifespan of these animals without reducing fecundity. Lifespan extension is dependent on functional Sir2, and is not observed when nutrients are restricted. Together these data indicate that STACs slow metazoan ageing by mechanisms that may be related to caloric restriction.

Sir2-like proteins (sirtuins) are a family of NAD⁺-dependent deacetylases conserved from *Escherichia coli* to humans^{5–9} (Fig. 1a) that play important roles in gene silencing, DNA repair, rDNA recombination and ageing in model organisms^{2,10–12}. When diet is restricted (caloric restriction), lifespan is extended in diverse species, suggesting that there is a conserved mechanism for nutrient regulation of ageing^{13–17}. In budding yeast, extra copies of SIR2 extend lifespan by 30%, apparently by mimicking caloric restriction^{1,18}. We recently described a group of compounds (STACs) that stimulate the catalytic activity of yeast and human sirtuins, and extend the replicative lifespan of yeast cells by up to 60% (ref. 4).

To establish whether STACs could activate sirtuins from multicellular animals, we developed a cell-based deacetylation assay for *D. melanogaster* S2 cells. Unlike other classes of deacetylases, the sirtuins are insensitive to the inhibitor trichostatin A (TSA). Several classes of polyphenolic STACs, including chalcones, flavones and stilbenes, increased the rate of TSA-insensitive deacetylation (Fig. 1b). To determine whether this activity was due to direct stimulation of a Sir2 homologue, we purified recombinant SIR-2.1 of *C. elegans* and Sir2 of *D. melanogaster* and determined the effect of various STACs on enzymatic activity *in vitro* (Fig. 1c, d). In a

dose-dependent manner, resveratrol stimulated deacetylation up to 2.5-fold for SIR-2.1 (Fig. 1e) and 2.4-fold for Sir2 (Fig. 1f). As previously observed with the yeast and human Sir2 enzymes, resveratrol lowered the K_m of SIR-2.1 for the co-substrate NAD⁺ (Fig. 1g).

Because resveratrol can significantly extend replicative lifespan in yeast⁴, we asked whether STACs could also extend lifespan in the metazoans *C. elegans* and *D. melanogaster*. Wild-type worms were transferred to plates containing 0 or 100 μ M of resveratrol shortly after reaching adulthood. Lifespan was extended up to 14%, using

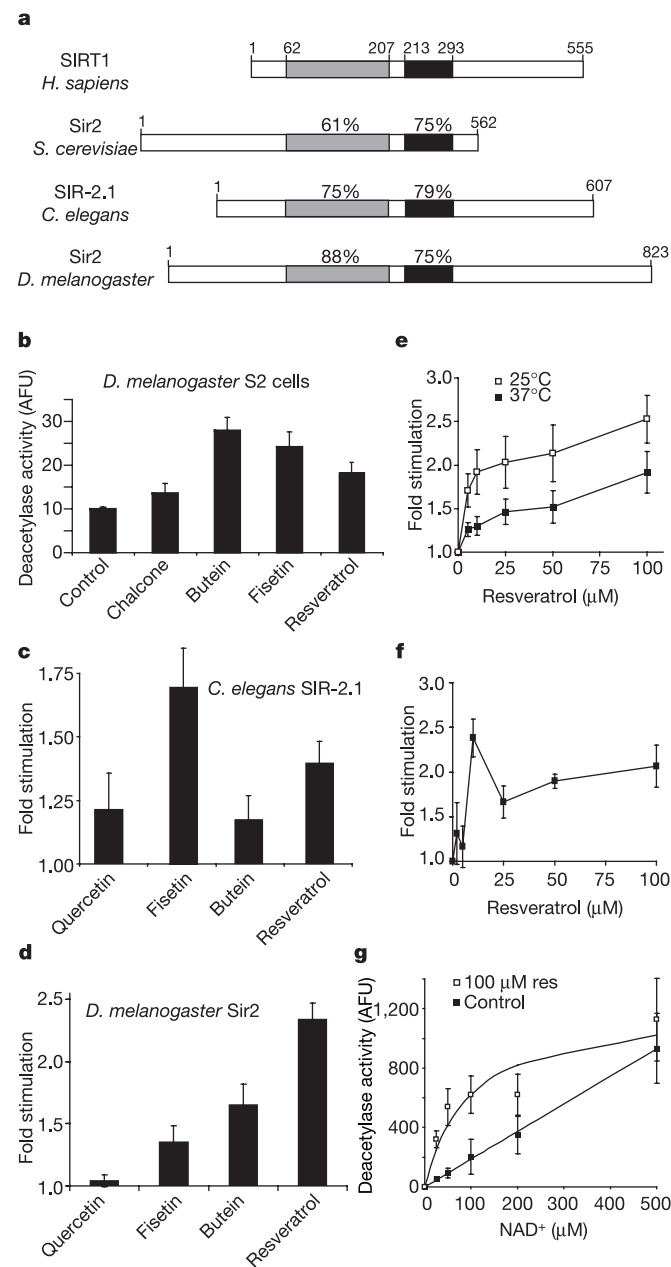


Figure 1 Effect of polyphenolic STACs on sirtuins. **a**, Sir2 polypeptides from various species. The NAD⁺-binding pocket (grey), substrate-binding groove (black) and per cent homology to SIRT1 are shown. **b**, Effect of polyphenolic STACs (500 μ M) on TSA-insensitive deacetylase activity in *Drosophila* S2 cells. **c**, Fold stimulation of SIR-2.1 by STACs (10 μ M). **d**, Fold stimulation of *Drosophila* Sir2 by STACs (10 μ M). Values are the mean of at least three determinations ($\pm$ s.e.). **e**, **f**, Activation of *C. elegans* SIR-2.1 (**e**) and *Drosophila* Sir2 (**f**) by resveratrol ($\pm$ s.e.). **g**, SIR-2.1 initial rate as a function of NAD⁺ concentration ($\pm$ s.e.). AFU, arbitrary fluorescence units.

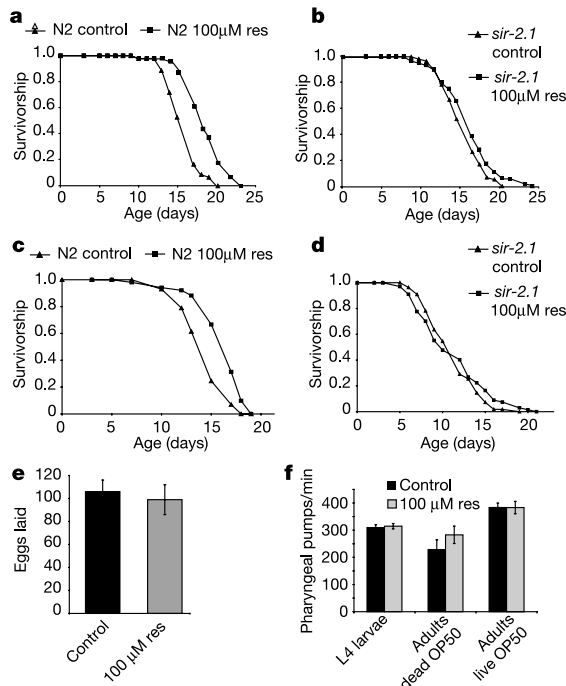


Figure 2 *C. elegans* survival on resveratrol. **a**, Survivorship of adult wild-type N2 *C. elegans* treated with resveratrol fed with heat-killed OP50 *E. coli*. **b**, Survivorship of *sir-2.1* mutants treated with resveratrol fed with heat-killed OP50. Adult lifespan of *sir-2.1* animals does not differ significantly from N2 controls. **c**, Survivorship of wild-type N2 *C. elegans* on 100 µM resveratrol fed with live OP50. **d**, Survivorship of *sir-2.1* mutants on 100 µM resveratrol fed with live OP50. **e**, Fecundity of adult hermaphrodites on 100 µM resveratrol ($\pm$ s.d.). **f**, Feeding rates of L4 larval and adult hermaphrodites treated with 100 µM resveratrol ($\pm$ s.d.).

either heat-killed or live *E. coli* as food supply (Fig. 2a, c, respectively; Supplementary Table ST1; averaged across trials at 100 µM resveratrol, lifespan was extended 10% on live and dead *E. coli*), and mortality was decreased across all adult ages (Supplementary Fig. S1). To test whether the lifespan extension depends on functional SIR-2.1, we constructed a *sir-2.1* null mutant. The lifespan of this strain was not appreciably shorter than the wild-type N2 control, and adults treated with resveratrol did not exhibit a significant lifespan extension relative to untreated worms (Fig. 2b, d; Supplementary Table ST1). There was no decrease in fecundity associated with resveratrol treatment (Fig. 2e). To rule out the possibility that resveratrol was causing the animals to eat less, thereby inducing a caloric restriction effect indirectly, we measured feeding rates of both L4 larval and adult worms with or without resveratrol and found no differences (Fig. 2f). Throughout the lifespan trials, we used the reproductive suppressant FUDR to prevent accumulation of progeny on the treatment plates.

We also tested whether STACs could extend lifespan in *D. melanogaster* using the standard laboratory wild-type strain Canton-S and normal fly culturing conditions (vials), and a *yw* marked wild-type strain and demographic culturing conditions (cages) (Supplementary Table ST2). Across independent tests in males and females on abundant diet, lifespan was extended up to 23% with fisetin and up to 29% with resveratrol (Fig. 3a, c, e; Supplementary Table ST2; averaged across these trials with 100 µM resveratrol, lifespan was extended 20% in females and 16% in males). Increased longevity was associated with reduced mortality before day 40 (Supplementary Fig. S1). A restricted diet increased lifespan by 40% in females and by 14% in males (averaged across trials), and under these conditions neither resveratrol nor fisetin further increased longevity (Fig. 3b, d, f). The lack of an additive effect of resveratrol and a low calorie diet suggests that resveratrol extends lifespan through a mechanism related to caloric restriction.

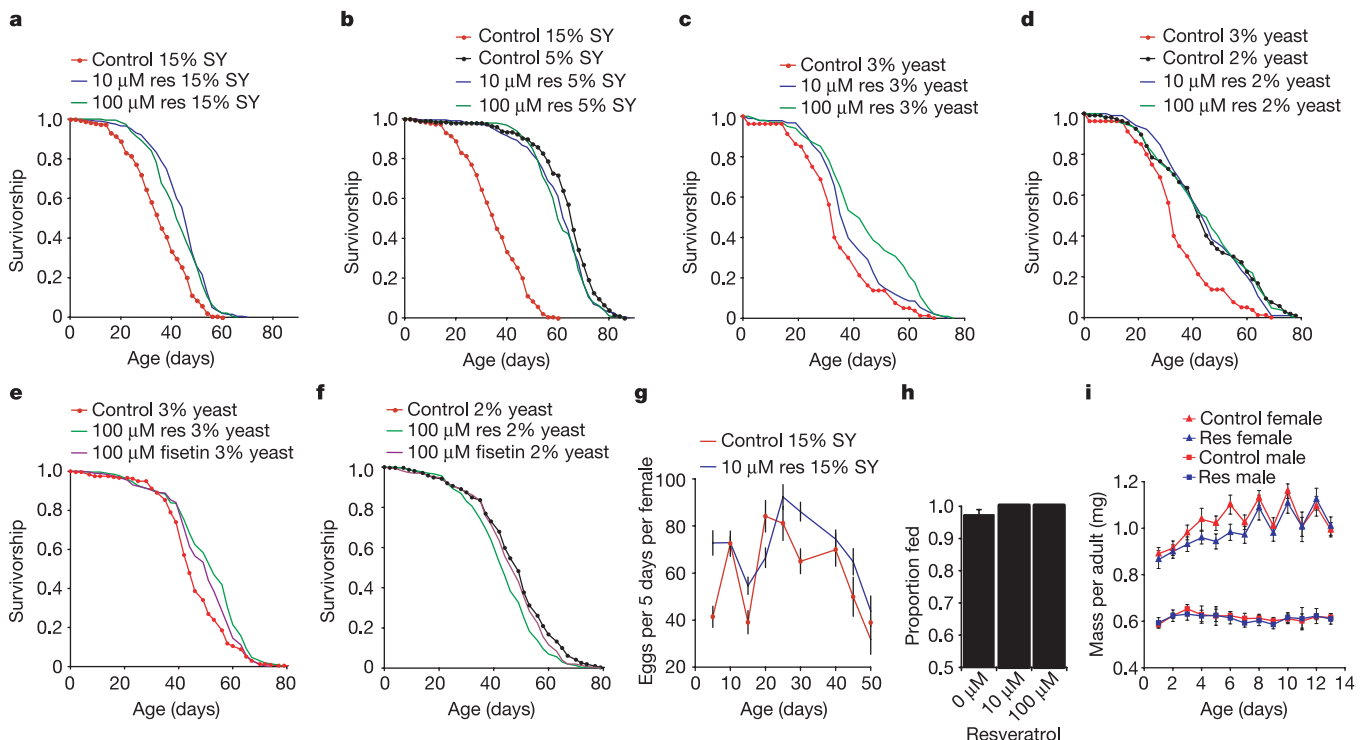


Figure 3 Survival of wild-type female *D. melanogaster* adults fed resveratrol or fisetin. **a**, Canton-S on 15% SY media. **b**, Canton-S on 5% SY media. **c**, *yw* on 3% CSY media. **d**, *yw* on 2% CSY media. **e**, *yw* on 3% CSY media. **f**, *yw* on 2% CSY media. **g**, Mean 5 day

fecundity per female of Canton-S on 15% SY media $\pm$ resveratrol. **h**, Proportion of *yw* females feeding on diet $\pm$ resveratrol in crop-filling assay. **i**, Mean body mass of Canton-S flies on diet $\pm$ resveratrol (10 µM). Errors represent s.e.

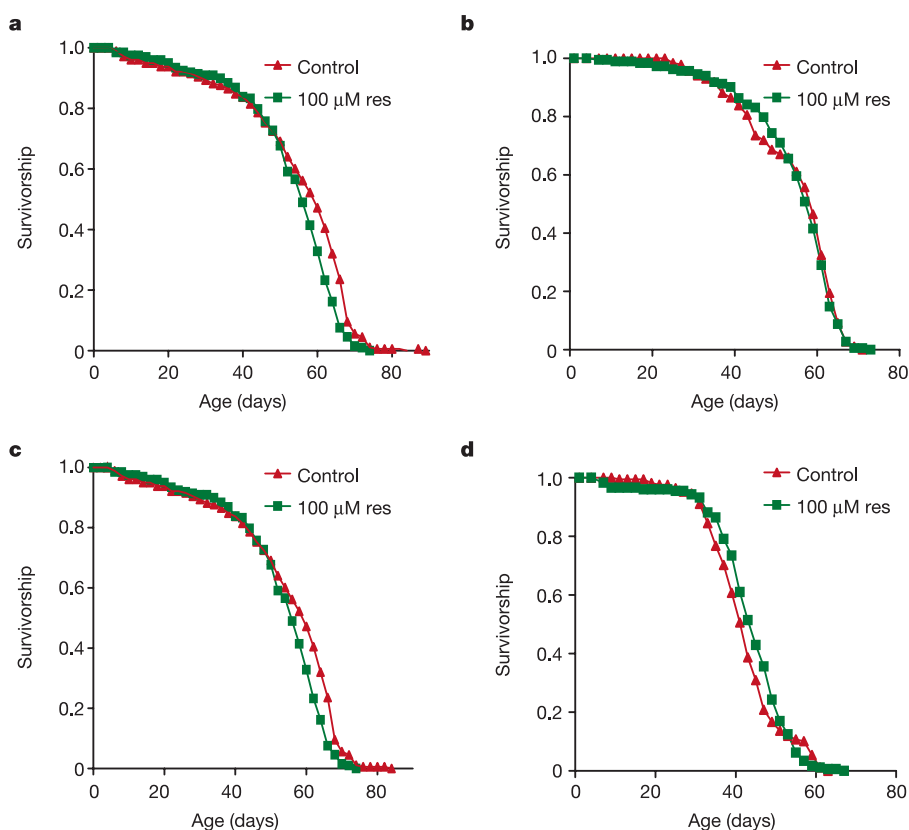


Figure 4 Survival of *D. melanogaster* adults with mutant alleles of *Sir2* fed resveratrol (100 μ M). Females (**a**) and males (**b**) with loss-of-function genotype *Sir2*^{4.5}/*Sir2*^{5.26}. Females (**c**) and males (**d**) with strong hypomorphic genotype *Sir2*¹⁷/*Sir2*^{KG00871}.

Surprisingly, while diet manipulations that extend *D. melanogaster* longevity typically reduce fecundity^{19,20}, longevity-extending doses of resveratrol modestly increased egg production (10 μ M resveratrol: 69.8 eggs/5 days, s.e. = 2.2; control: 59.9 eggs/5 days, s.e. = 2.2; $t = 3.17$, $P = 0.0017$), particularly in the earliest days of adult life (Fig. 3g). The increase in egg production suggests that the lifespan extending effect of resveratrol in *D. melanogaster* was not due to caloric restriction induced by food aversion or lack of appetite. Consistent with this, no decrease in food uptake was seen with resveratrol-fed flies (Fig. 3h). Furthermore, resveratrol-fed flies maintained normal weight (Fig. 3i), except during days 3–6 when resveratrol fed females were laying significantly more eggs than control fed females.

To determine whether resveratrol extends fly lifespan in a *Sir2*-dependent manner, we analysed a *Sir2* allelic series with increasing amounts of *Sir2*. Adult offspring from crosses between independently derived alleles of *Sir2* were tested. Resveratrol failed to extend lifespan in flies completely lacking functional *Sir2* (*Sir2*^{4.5}/*Sir2*^{5.26}) (Fig. 4a, b) or in flies in which *Sir2* is severely decreased (*Sir2*¹⁷/*Sir2*^{KG00871}) (Fig. 4c, d). Resveratrol increased longevity a small but statistically significant amount in flies homozygous for a hypomorphic *Sir2* allele (*Sir2*^{KG0087}/*Sir2*^{KG0087}) (Supplementary Table ST2, trial 6) and increased lifespan up to 17% in flies with one copy of the hypomorphic allele and one copy of a wild-type *Sir2* (Canton-S/*Sir2*^{KG0087}) (Supplementary Table ST2, trial 7). These data suggest that the ability of resveratrol to extend fly lifespan requires functional *Sir2*.

We previously reported that STACs extend the replicative lifespan of yeast cells by mimicking caloric restriction⁴. Here we show that STACs can extend lifespan in *C. elegans* and *D. melanogaster*, both of which are composed of primarily non-dividing (post-mitotic) cells as adults, and whose somatic and reproductive ageing are indepen-

dent measures of senescence. In both species, resveratrol increases lifespan in a *Sir2*-dependent manner and, at least for the fly, this action appears to function through a pathway related to caloric restriction.

Our observation that resveratrol can increase longevity without an apparent cost of reproduction is counter to prevalent concepts of senescence evolution. However, STACs may still entail trade-offs that involve other traits, or that occur only under some environmental conditions^{21,22}. Plants synthesize STACs such as resveratrol in response to stress and nutrient limitation²³, possibly to activate their own sirtuin pathways⁴. These molecules may activate animal sirtuins because they serve as plant defence mechanisms against consumers or because they are ancestrally orthologous to endogenous activators within metazoans. Alternatively, animals may use plant stress molecules as a cue to prepare for a decline in their environment or food supply⁴. Understanding the adaptive significance, endogenous function and evolutionary origin of sirtuin activators will lead to further insights into the underlying mechanisms of longevity regulation, and could aid in the development of interventions that provide the health benefits of caloric restriction. □

Methods

C. elegans media, strains, lifespan and feeding assays

Bristol N2 (*Caenorhabditis Genetics Center*) was used as the wild-type strain. The *sir-2.1* mutant strain was generated by backcrossing VC199 (*sir-2.1(ok434)*) to N2 four times. For the lifespan assays, synchronized animals were transferred to treatment plates as young adults (2 days after hatching, day 0 of assay), and were transferred to fresh treatment plates every 2 days for the first 6 to 8 days of the assay. Treatment plates were standard NGM media with the reproductive suppressant FUDR (Sigma; 100 μ g l⁻¹) containing resveratrol or solvent (DMSO, which does not affect lifespan) added either directly into the agar before pouring (for live OP50 trials) or diluted into PBS and added to the surface of a dry plate to the indicated final concentration (for dead OP50 trials). For heat-killed *E. coli* trials, OP50 cultures were heated to 65 °C for 30 min, then pelleted and resuspended in

1/10 volume in S Basal supplemented with 10 mM MgSO₄. For *E. coli* killed by ampicillin and kanamycin, overnight cultures of OP50 were treated with 500 µg ml⁻¹ ampicillin and 250 µg ml⁻¹ kanamycin overnight at 37 °C. In all assays, worms were monitored daily for mortality by gently probing with a platinum pick. Assays were performed at 24 °C.

To assay worm feeding rates, worms at the indicated stages were placed on treatment plates (no FUDR) for 4–5 h, then videoed for 1 min using a Pixelink PL-662 camera. The frame rate was slowed and the pumping rate of the pharynx was counted. To assay fecundity, gravid hermaphrodites (5 per plate, raised from synchronized L1s on normal or treatment plates) were allowed to lay eggs on their respective media for 5 h, and the total number of eggs was counted.

D. melanogaster media, strains, feeding assay and lifespan assays

Survival assays were conducted independently with adult *D. melanogaster* in two laboratories. At Brown University, all trials used an *yw* marked wild-type strain. Larvae were reared on standard cornmeal-sugar-yeast (CSY) agar diet (cornmeal 5%, sucrose 10.5%, SAF yeast 2% and agar 0.7%). Newly eclosed adults were placed in 1-litre demography cages with approximately 75 males and 75 females. Three to four replicate 1-litre demography cages were used for each treatment group in each trial. Every two days, dead flies were removed and scored, and food vials were replenished. Food vials contained CSY diet with SAF yeast as either 2% or 3% by weight. Test compounds in 100 µl of EtOH (or blank EtOH in controls) were mixed into melted aliquots of the adult food media to make a final concentration of 0, 10 or 100 µM. Fresh stock solutions and adult media were prepared weekly. At the University of Connecticut Health Center, lifespan trials were conducted with the wild type strain Canton-S, *Sir2*^{4,5} and *Sir2*^{5,26} (S. Smolik), *Sir2*²⁷ (S. Astrom) and *Sir2*^{KG00871} (Drosophila Stock Center). Larvae for all tests were reared on standard CSY diet. Newly eclosed adults were incubated in plastic shell vials containing 5 ml of 15% sugar-yeast diet (15% SY) or 5% sugar-yeast (5% SY) diet (15% SY: 15% yeast, 15% sucrose, 2% agar; 5% SY: 5% yeast, 5% sucrose, 2% agar as per ref. 20.). In all trials, ~20 males with ~20 females were placed into each of 10 vials per treatment group. Every two days, flies were passed into new vials and dead flies were counted. Resveratrol in EtOH (or EtOH alone in controls) was added to the media during its preparation after it had cooled to 65 °C and mixed vigorously. Final compound concentrations were 0, 10, 100 or 200 µM. Fresh stock solution and adult media was prepared weekly.

Feeding rate was measured in *yw* females with the crop-filling assay²⁴. Females were held overnight with water and placed on 2% CSY diet containing food colour (FDA Blue 1) and 0, 10 or 100 µM resveratrol with EtOH. The presence of dye-marked food in the crop was scored in sets of 20 females across five 5-min intervals. For body mass measurements, 10 vials with 20 males and 20 females each of wild type CS-5 flies were kept on 15% SY diet with EtOH or with resveratrol in EtOH (10 µM). Males and females were weighed daily.

Sirtuin purification and deacetylation assays

These are described in Supplementary Methods.

Received 13 April; accepted 28 June 2004; doi:10.1038/nature02789.

Published online 14 July 2004.

1. Lin, S. J., Defossez, P. A. & Guarente, L. Requirement of NAD and SIR2 for life-span extension by calorie restriction in *Saccharomyces cerevisiae*. *Science* **289**, 2126–2128 (2000).
2. Gasser, S. C. M. The molecular biology of the SIR proteins. *Gene* **279**, 1–16 (2001).
3. Hekimi, S. & Guarente, L. Genetics and the specificity of the aging process. *Science* **299**, 1351–1354 (2003).
4. Howitz, K. T. *et al.* Small molecule activators of sirtuins extend *Saccharomyces cerevisiae* lifespan. *Nature* **425**, 191–196 (2003).
5. Landry, J. *et al.* The silencing protein SIR2 and its homologs are NAD-dependent protein deacetylases. *Proc. Natl Acad. Sci. USA* **97**, 5807–5811 (2000).
6. Imai, S., Armstrong, C. M., Kaerberlein, M. & Guarente, L. Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase. *Nature* **403**, 795–800 (2000).
7. Smith, J. S. *et al.* A phylogenetically conserved NAD⁺-dependent protein deacetylase activity in the SIR2 protein family. *Proc. Natl Acad. Sci. USA* **97**, 6658–6663 (2000).
8. Tanner, K. G., Landry, J., Sternglanz, R. & Denu, J. M. Silent information regulator 2 family of NAD-dependent histone/protein deacetylases generates a unique product, 1-O-acetyl-ADP-ribose. *Proc. Natl Acad. Sci. USA* **97**, 14178–14182 (2000).
9. Tanny, J. C., Dowd, G. J., Huang, J., Hilz, H. & Moazed, D. An enzymatic activity in the yeast Sir2 protein that is essential for gene silencing. *Cell* **99**, 735–745 (1999).
10. Guarente, L. Sir2 links chromatin silencing, metabolism, and aging. *Genes Dev.* **14**, 1021–1026 (2000).
11. Tissenbaum, H. A. & Guarente, L. Increased dosage of a sir-2 gene extends lifespan in *Caenorhabditis elegans*. *Nature* **410**, 227–230 (2001).
12. Rogina, B., Helfand, S. L. & Frankel, S. Longevity regulation by *Drosophila* Rpd3 deacetylase and caloric restriction. *Science* **298**, 1745 (2002).
13. Jiang, J. C., Jaruga, E., Repnevskaya, M. V. & Jazwinski, S. M. An intervention resembling caloric restriction prolongs life span and retards aging in yeast. *FASEB J.* **14**, 2135–2137 (2000).
14. Kenyon, C. A conserved regulatory mechanism for aging. *Cell* **105**, 165–168 (2001).
15. Masoro, E. J. Caloric restriction and aging: an update. *Exp. Gerontol.* **35**, 299–305 (2000).
16. Koubova, J. & Guarente, L. How does calorie restriction work? *Genes Dev.* **17**, 313–321 (2003).
17. Sinclair, D. A. Paradigms and pitfalls of yeast longevity research. *Mech. Ageing Dev.* **123**, 857–867 (2002).
18. Kaerberlein, M., McVey, M. & Guarente, L. The SIR2/3/4 complex and SIR2 alone promote longevity in *Saccharomyces cerevisiae* by two different mechanisms. *Genes Dev.* **13**, 2570–2580 (1999).
19. Chippindale, A. K., Leroy, A. M., Kim, S. B. & Rose, M. R. Phenotypic plasticity and selection in *Drosophila* life-history evolution. *J. Evol. Biol.* **6**, 171–193 (1993).
20. Chapman, T. & Partridge, L. Female fitness in *Drosophila melanogaster*: an interaction between the effect of nutrition and of encounter rate with males. *Proc. R. Soc. Lond. B* **263**, 755–759 (1996).
21. Walker, D. W., McColl, G., Jenkins, N. L., Harris, J. & Lithgow, G. J. Evolution of lifespan in *C. elegans*. *Nature* **405**, 296–297 (2000).
22. Marden, J. H., Rogina, B., Montooth, K. L. & Helfand, S. L. Conditional tradeoffs between aging and

organismal performance of Indy long-lived mutant flies. *Proc. Natl Acad. Sci. USA* **100**, 3369–3373 (2003).

23. Soles, G. J., Diamandis, E. P. & Goldberg, D. M. Resveratrol: a molecule whose time has come? And gone? *Clin. Biochem.* **30**, 91–113 (1997).
24. Edgecomb, R. S., Harth, C. E. & Schneiderman, A. M. Regulation of feeding behavior in adult *Drosophila melanogaster* varies with feeding regime and nutritional state. *J. Exp. Biol.* **197**, 215–235 (1994).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank S. Parkhurst for the dSir2-pRSETc plasmid, and J. Whetstone, H. Yang and Y. Shi for advice and reagents. *C. elegans* strain VC199 (*sir-2.1(ok434)*) was generated by the *C. elegans* Reverse Genetics Core Facility at the University of British Columbia. This work was supported by the National Institute on Aging, the Harvard-Armstrong Foundation, the Donaghy Foundation, and a generous gift from Harmon Rasnow. S.L.H. is an Ellison Medical Research Foundation Senior Investigator, and D.S. and M.T. are Ellison Medical Research Foundation Fellows. J.W. is supported by an NSF Graduate Research Fellowship.

Competing interests statement The authors declare competing financial interests: details accompany the paper on www.nature.com/nature.

Correspondence and requests for materials should be addressed to D.S. (david_sinclair@hms.harvard.edu), S.L.H. (shelfand@neuron.uchc.edu) or M.T. (marc_tatar@brown.edu).

Planar cell polarity signalling controls cell division orientation during zebrafish gastrulation

Ying Gong^{1,2}, Chunhui Mo² & Scott E. Fraser^{1,2}

¹Biological Imaging Center, Beckman Institute, ²Biochemistry and Molecular Biophysics, Division of Biology, California Institute of Technology, Pasadena, California 91125, USA

Oriented cell division is an integral part of pattern development in processes ranging from asymmetric segregation of cell-fate determinants to the shaping of tissues^{1,2}. Despite proposals that it has an important function in tissue elongation^{3,4}, the mechanisms regulating division orientation have been little studied outside of the invertebrates *Caenorhabditis elegans* and *Drosophila melanogaster*¹. Here, we have analysed mitotic divisions during zebrafish gastrulation using *in vivo* confocal imaging and found that cells in dorsal tissues preferentially divide along the animal–vegetal axis of the embryo. Establishment of this animal–vegetal polarity requires the Wnt pathway components Silberblick/Wnt11, Dishevelled and Strabismus. Our findings demonstrate an important role for non-canonical Wnt signalling in oriented cell division during zebrafish gastrulation, and indicate that oriented cell division is a driving force for axis elongation. Furthermore, we propose that non-canonical Wnt signalling has a conserved role in vertebrate axis elongation, orienting both cell intercalation and mitotic division.

Gastrulation in zebrafish starts as the blastoderm begins to cover the yolk cell. Mesendoderm precursors internalize near the blastoderm margin to form an inner blastoderm stratum termed the hypoblast. Cells remaining in the outer stratum constitute the epiblast, which give rise to neural ectoderm on the dorsal side and epidermis on the ventral side^{5,6}. The dorsal epiblast consists of two–three layers of cells, most of which divide twice during gastrulation: once near the start of gastrulation and another in mid gastrulation^{7,8}. Mitotic divisions at different stages appear to be random, oriented along the animal–vegetal axis, or oriented medio-laterally^{7,8}. Neither the full extent nor the mechanisms that regulate division orientation have been explored. Thus, we have characterized the patterns of cell division throughout the depth of the dorsal

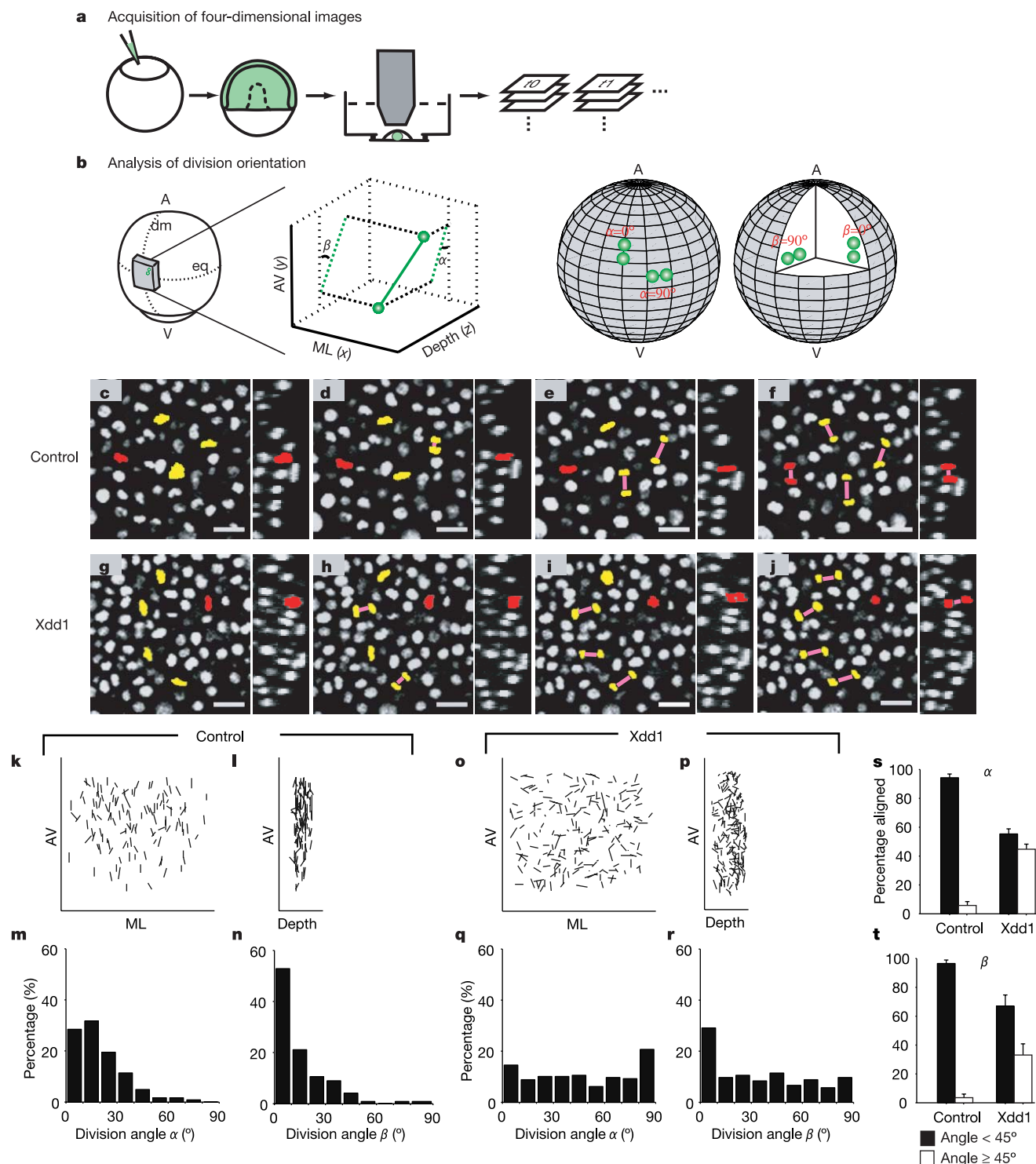


Figure 1 Imaging analysis of division orientation in zebrafish gastrula. **a**, Embryos are injected with various constructs, raised to early gastrula stage, and cultured on the microscope for four-dimensional imaging. **b**, Quantitative analysis of division orientation. A pair of daughter nuclei is shown as green spheres. The two panels on the right are schematic diagrams showing α and β within the whole embryo. dm, dorsal midline; eq, equator. **c–j**, Confocal time-lapse images from a representative control (**c–f**) and Xdd1-expressing (**g–j**) embryo. Main panels are AV–ML, whereas smaller right-hand panels are AV–depth cross-sections, respectively. Dividing nuclei are highlighted in yellow (shown in AV–ML only) or red (shown in both AV–ML and AV–depth), and division axes are marked by magenta lines. Divisions are aligned with the AV axis in the control but

randomized in the Xdd1-expressing embryo. Animal pole is up in these images. Scale bar, 20 μ m; time interval, 2 min. **k, l, o, p**, In control (**k, l**) and Xdd1-expressing embryos (**o, p**), each division is represented by a line. Most lines are AV oriented in the control but randomly oriented in the Xdd1-expressing embryo. **m, n, q, r**, Histograms showing the distribution of α and β angles in control (**m, n**) and Xdd1-expressing (**q, r**) embryos. Most angles are close to 0° in the control embryo but are distributed nearly uniformly from 0 to 90° in the Xdd1-expressing embryo. **s, t**, Bar plots of division angles less than and greater than 45° (mean $\pm$ s.d.; control, 521 divisions from 4 embryos; Xdd1, 533 divisions from 3 embryos).

epiblast during zebrafish gastrulation and have investigated its molecular control.

To monitor mitotic divisions through the depth of the epiblast, we imaged zebrafish gastrulae ubiquitously expressing histone 2B fused to green fluorescent protein (H2B–GFP) using four-dimensional (x, y, z axes over time (t)) confocal microscopy (Fig. 1a). H2B–GFP marks the chromatin and renders mitosis easily visible in time-lapse movies owing to the changes in chromosome arrangement during mitosis. Division orientation is determined from the four-dimensional images by measuring the angle between the mitotic spindle and the animal–vegetal (AV) axis (Fig. 1b). Each angle is characterized by its two planar projections: α , the angle within the epiblast plane, and β , the angle within the perpendicular plane. Analysis of wild-type gastrulae showed that divisions in all layers of the dorsal epiblast are α -oriented along the AV axis (Fig. 1c–f, k–n). Over 90% of α angles are under 45° , with 58% under 20° (Fig. 1s). Similarly, 95% of β angles are under 45° , with 62% under 20° (Fig. 1t).

Previous observations have shown that the mitotic spindle tends to align with the cell's long axis owing to geometric constraints (known as Hertwig's rule)^{9,10}. To investigate whether this is the case in our system, we examined the relationship between the division orientation of dorsal epiblast cells and their morphology by double-labelling the cells with a membrane red fluorescent protein (RFP) and H2B–GFP. In our time-lapse movies, most (92%) of the dorsal epiblast cells in wild-type embryos elongate with a mediolateral (ML) bias at interphase (Fig. 2a; green circles in c; length to width ratio = 1.66 ± 0.45). During mitotic division, these cells round up and divide to create daughter cell pairs that separate in the AV orientation (Fig. 2a, d). Thus, perhaps not surprisingly, in 65% of the cases the angle between a cell's long axis during interphase and its spindle during mitosis is within 10° of perpendicular (Fig. 2e). Thus, most cells divide nearly orthogonal to their long axis, the opposite of the expectation from Hertwig's rule.

The above findings suggest that the mitotic spindle is actively

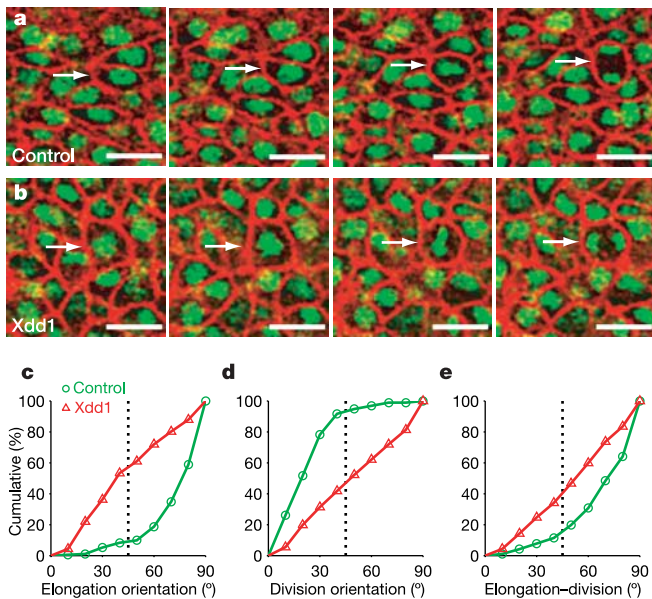


Figure 2 Relationship between cell elongation and division orientations. **a, b**, Confocal time-lapse images from a typical control (**a**) and Xdd1-expressing (**b**) embryo double-labelled with membrane RFP (red) and H2B–GFP (green). Dividing cells are marked by arrows. Scale bar, $20 \mu\text{m}$; time interval, 2 min. **c**, Cumulative plot of cell elongation orientation with respect to AV. Control cells elongate mediolaterally, whereas Xdd1-expressing cells elongate in random orientations. **d**, Cumulative plot of cell division orientation with respect to AV, showing randomized orientation caused by Xdd1. **e**, Cumulative plot of the angle between a cell's elongation axis and division axis. Control, 192 cells from 3 embryos; Xdd1, 184 cells from 3 embryos.

oriented in the dorsal epiblast, and prompted us to investigate the molecular mechanisms underlying this process. Signalling through Dishevelled (Dsh) has been implicated in polarized cell division in *Drosophila* and *C. elegans*¹, raising the possibility that it regulates oriented division in zebrafish gastrulae.

To assess the function of Dsh, we injected into zebrafish embryos messenger RNA encoding Xdd1, a mutant form of *Xenopus* Dsh that blocks axis elongation in *Xenopus* and zebrafish (Fig. 3b)^{11–13}. Our analysis of division orientation showed that Xdd1 severely disrupts

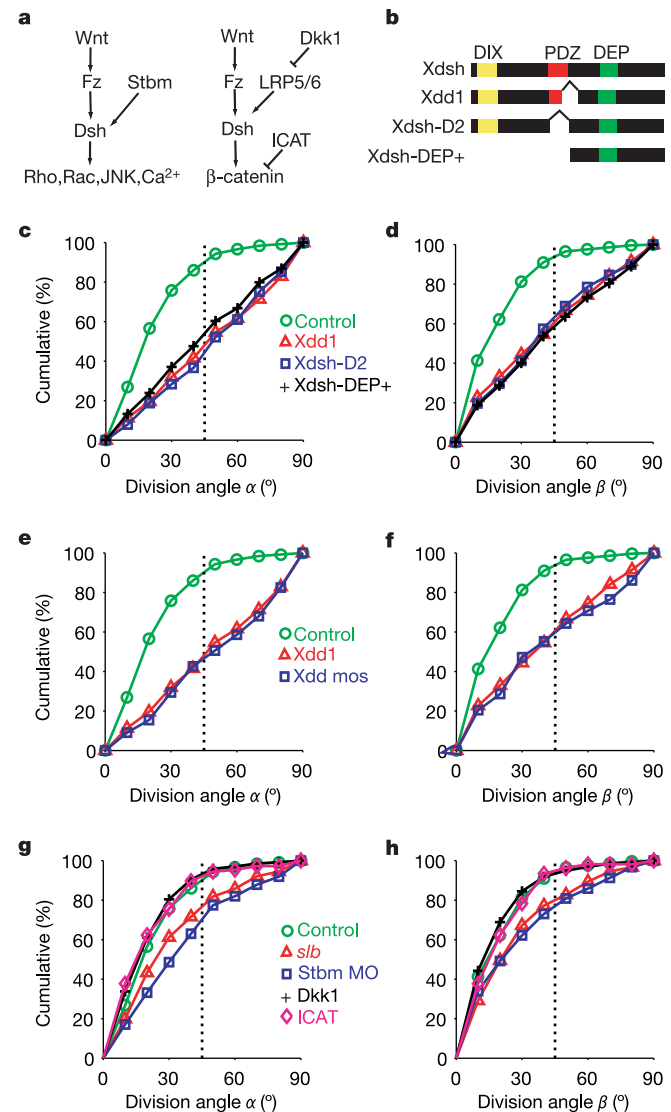


Figure 3 Components of the PCP pathway regulate division orientation. **a**, Diagram of the canonical Wnt/ β -catenin (right panel) and non-canonical Wnt/PCP (left panel) pathways. Fz, Frizzled. **b**, Dsh deletion constructs tested. **c, d**, Cumulative plots of division angles from embryos expressing various Dsh constructs. Xdd1, Xddsh-D2 and Xddsh-DEP+ disrupt division orientation polarity to a similar degree. Control, 521 divisions from 4 embryos; Xdd1, 533 divisions from 3 embryos; Xddsh-D2, 378 divisions from 2 embryos; Xddsh-DEP+, 555 divisions from 3 embryos. **e, f**, Xdd1 affects division cell autonomously. Xdd1 mos, Xdd1 mosaic clones in wild-type backgrounds; 138 divisions from 6 embryos. The wild-type and Xdd1 curves are plotted for comparison. **g, h**, Cumulative plots of division angles from embryos in which different canonical Wnt or PCP pathway components are compromised. Mutation in the *slb/wnt11* gene and knockdown of Stbm by a morpholino (Stbm MO) both randomize division orientations. Neither Dkk1 nor ICAT affect the orientation distribution significantly. *slb*, 411 divisions from 3 embryos; Stbm MO, 638 divisions from 4 embryos; Dkk1, 500 divisions from 5 embryos; ICAT, 106 divisions from 6 embryos. The dotted lines represent 45° thresholds.

the AV alignment of divisions in all layers of the epiblast (Fig. 1). Cells now divide in a randomized fashion, with spindle orientation almost uniformly distributed from 0 to 90° (Compare Fig. 1o, p with k, l and q, r with m, n). These data are summarized in Fig. 1s, t. In *Xdd1*-expressing embryos, only about one-half (55%) of all α angles are within 45°, compared with over 90% in the control (Fig. 1s). Similarly, 65% of all β angles are within 45° in *Xdd1*-expressing embryos, whereas the number is over 95% in the control (Fig. 1r). In addition to randomizing division orientation of dorsal epiblast cells, *Xdd1* also disrupts the polarity of elongation in these cells (length to width ratio = 1.48 ± 0.33 ; Fig. 2b, red triangles in c). Furthermore, cells divide at random angles rather than perpendicularly to their long axis (Fig. 2e).

Xdd1 disrupts convergence and extension of the dorsal tissue (Fig. 4b)^{11,13}. It is thus possible that the disruption of oriented division observed in *Xdd1*-overexpressing embryos was due to compromised morphogenesis of the tissue. To address this, we generated mosaic clones of *Xdd1*-expressing cells in a wild-type background by injecting a single cleavage-stage blastomere. Such embryos undergo normal morphogenesis, and are indistinguishable from unmanipulated controls morphologically (data not shown). Subsequent analysis shows that these mosaic *Xdd1*-expressing cells have randomized division orientation, with angular distribution similar to that of embryos overexpressing *Xdd1* ubiquitously (Fig. 3e, f; see Supplementary Information for details). Thus, Dsh has a cell-autonomous function and exerts its effect on division orientation directly.

Dsh is involved in multiple Wnt pathways including the canonical Wnt/ β -catenin pathway and the non-canonical pathway related to the *Drosophila* planar cell polarity (PCP) pathway (Fig. 3a). *Xdd1* is a strong inhibitor of PCP signalling, but can block canonical Wnt/ β -catenin signalling in a *Xenopus* secondary axis assay¹². To distinguish

between these two pathways, we tested two additional *Xdsh* constructs: *Xdsh*-D2 and *Xdsh*-DEP+ (Fig. 3b)^{11,12,14}. *Xdsh*-D2, which lacks the entire PDZ domain, strongly inhibits PCP but is fully functional for canonical signalling^{11,14}. *Xdsh*-DEP+, which only has the DEP domain and lacks the DIX and PDZ domains, blocks PCP and is not functional for canonical signalling¹².

Analysis of cell divisions in zebrafish embryos expressing either *Xdsh*-D2 or *Xdsh*-DEP+ revealed that division orientation is randomized to a similar extent as caused by *Xdd1* (Fig. 3c, d). Therefore, although the three different Dsh deletion constructs can have very different effects on canonical Wnt signalling, they have equivalent ability to inhibit division alignment. This suggests that Dsh controls division orientation in the dorsal epiblast independently of the canonical Wnt signalling pathway.

To dissect further the signalling pathways involved, we functionally blocked the canonical Wnt pathway with two specific antagonists: Dickkopf-1 (*Dkk1*)¹⁵ and inhibitor of β -catenin and TCF-4 (ICAT)¹⁶. *Dkk1* antagonizes Wnt signalling by binding to lipoprotein-receptor-related protein 5/6 (LRP5/6), a possible Wnt co-receptor that is specifically required for canonical Wnt/ β -catenin signalling (Fig. 3a)¹⁷. Whereas embryos overexpressing the Dsh constructs exhibit mediolaterally expanded somites, embryos overexpressing zebrafish *Dkk1* exhibit a visibly different phenotype with diminished somites and enlarged head (data not shown)¹⁵, suggesting that the major effect of the Dsh constructs is different from that of *Dkk1*. Notably, *Dkk1* does not affect the AV alignment of divisions (Fig. 3g, h; note the overlap of the control and *Dkk1* curves). A similar result is obtained using ICAT, a protein that binds to β -catenin and inhibits its transcriptional function in the canonical Wnt pathway (Fig. 3a). Because uniform overexpression of ICAT during early development interferes with the activity of the Nieuwkoop centre mediated by the maternal canonical Wnt pathway and causes gross dorsal–ventral patterning defects, we generated mosaic clones of ICAT-overexpressing cells on the dorsal side. The distribution of division orientations in these cells is very similar to that in control cells (Fig. 3g, h). Together, these results show that the non-canonical Wnt pathway is not involved directly in regulating division orientation during gastrulation.

To confirm that PCP signalling regulates oriented cell division in zebrafish, we tested the function of other factors implicated in PCP (that is, Wnt11 and Strabismus (*Stbm*); Fig. 3a). Wnt11 is a ligand for the PCP pathway in zebrafish and *Xenopus*^{12,18}. In our *in vivo* time-lapse analysis of *silberblick* (*slb*)/*wnt11* loss-of-function mutants, mitotic divisions in the dorsal epiblast cells exhibit less pronounced AV polarity relative to control cells in both the epiblast and perpendicular planes (Fig. 3g, h; compare red triangles with green circles). The orientation of division in *slb* mutant embryos is significantly less disrupted than embryos injected with the Dsh reagents (compare Fig. 3g, h with c, d), suggesting that additional factors signal through Dsh to regulate division orientation in the dorsal epiblast. Consistent with this notion, Wnt5a, another non-canonical Wnt, appears to act in parallel with Wnt11 in both zebrafish and *Xenopus*^{19,20}. The function of the PCP pathway was further tested by disrupting *Stbm*, a transmembrane protein that modulates PCP but does not lie in a linear cascade with Wnt/Dsh (Fig. 3a)^{13,21}. After injection with a *Stbm* morpholino²¹, mitotic divisions in the dorsal epiblast become misaligned (blue squares in Fig. 3g, h), revealing an important role for *Stbm* in this process.

Our experiments on the zebrafish dorsal epiblast show both a matching of division orientation and axis elongation normally, and a disruption of oriented cell division and axis elongation after inhibition of PCP signalling. We assessed the contribution of oriented cell division to axis elongation. Control embryos are well extended by the end of gastrulation, whereas the *Xdd1*-expressing embryos have a shorter axis, reaching only 56% of the length of the control (Fig. 4a, b, e). At the cellular level, we measured the change in AV dimension that resulted from one round of cell divisions

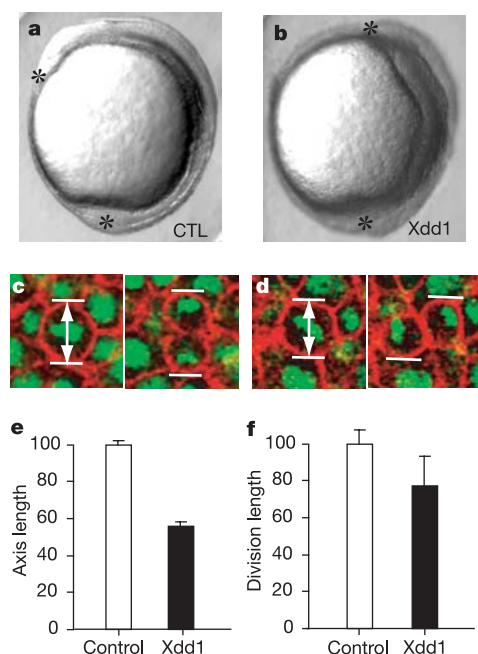


Figure 4 Contribution of oriented cell division to axis elongation. **a, b**, A representative *Xdd1*-expressing embryo has a shorter anterior–posterior axis (**b**) than the control (**a**) after gastrulation. Asterisks mark the anterior and posterior limits of the body axis. **c, d**, The AV dimension of a daughter cell pair after cytokinesis (marked by white lines) was measured and normalized by dividing the AV dimension of the mother cell (white arrows) at metaphase. Control (**c**) and *Xdd1*-expressing cells (**d**) are shown. **e**, Relative extension of the body axis in *Xdd1* embryos ($n = 11$) compared with controls ($n = 11$). **f**, Relative extension resulted from one round of cell divisions in *Xdd1*-expressing embryos ($n = 41$ cells from 2 embryos) compared with controls ($n = 46$ cells from 2 embryos). All values are mean $\pm$ s.d.

(Fig. 4c, d). Oriented divisions in control embryos increase the AV dimension by 1.5-fold, whereas randomized divisions in Xdd1-expressing embryos only increase it by 1.2-fold, which is 77% of the control level (Fig. 4f). This result indicates that oriented cell division accounts for a significant amount (but not total) of extension of the zebrafish gastrulae. Mediolateral intercalation has been shown to be a driving force for axis elongation in zebrafish and *Xenopus*^{22,23}. We propose that oriented cell division and cell intercalation, both under the regulation of the PCP pathway, collaborate to elongate the anterior–posterior body axis in zebrafish. Such a collaborative mechanism may also be used in axis elongation in the amniotes^{3,24}. We speculate that PCP signalling polarizes the cell, orienting both mediolateral intercalation and division axis.

Our results demonstrate the importance of non-canonical Wnt/PCP signalling in controlling cell division orientation and axis elongation in the early zebrafish embryo. Given that the PCP pathway has been shown to regulate spindle orientation during asymmetric cell division in *C. elegans* and *Drosophila*^{25,26}, our findings suggest that it has an evolutionarily conserved role in vertebrates. Oriented cell division is a common feature of many vertebrate developmental processes, ranging from the generation of cell layers²⁷ to primitive streak extension⁴ and neurogenesis²⁸. An interesting and testable possibility is that these processes have a similar dependence on PCP signalling. Together with our existing knowledge of the various roles of PCP signalling, our results highlight the central nature of the PCP pathway in regulating the multiple mechanisms that generate the elongation of the vertebrate embryo in its early development. □

Methods

mRNA and morpholino injection

Capped mRNA was injected into one-cell-stage embryos. The following dosages were used: Xdd1 (also known as Xdsh-ΔPDZ), 500 pg; Xdsh-DEP +, 800 pg; Xdsh-D2-GFP, 500 pg; Zdkk1, 500 pg; β-galactosidase, 500 pg (control). For cell division experiments, each mRNA was co-injected with 100 pg H2B-GFP²⁹ or H2B-RFP³⁰ mRNA. For Xdd1 mosaic experiments, a single cell at the 16–32-cell stage was injected with approximately 40 pg Xdd1 and 10 pg H2B-GFP mRNA. For ICAT mosaic experiments, a single non-marginal cell at the 64-cell stage was injected with approximately 40 pg of the mouse ICAT mRNA and 10 pg H2B-GFP mRNA. For double labelling experiments, 200 pg membrane RFP was also injected. Morpholino for the zebrafish Stbm mRNA has been described previously²¹. Two nanograms of morpholino was co-injected with 100 pg H2B-GFP into one-cell-stage embryos.

In vivo imaging

Embryonic shield-stage embryos were dechorionated and mounted in 1.3% low-gelling-point agarose in a custom-made imaging chamber. Four-dimensional confocal time-lapse imaging was carried out on an upright Zeiss LSM 510 or an inverted LSM PASCAL laser-scanning microscope. z-stacks were collected at 1.5–2-min intervals and spacing between consecutive z slices was set to values recommended by the microscope-controlling software (typically around 2 μm). Embryos were imaged for 3.5–4 h to approximately bud stage. The 488-nm laser line was used for imaging GFP and 543-nm laser line for RFP. Low laser power was used to minimize phototoxicity. Development of each embryo was monitored after the imaging session and embryos showing abnormal necrosis were not included for analysis.

Analysis of cell division orientation

Four-dimensional images were visualized using the Zeiss LSM software. Cell divisions (division 15, which occurs after the embryonic shield stage⁶) were easily identified in cells labelled with H2B-GFP because of changes in chromatin arrangement during mitosis. For each mitotic division, *x*, *y*, *z* coordinates of the two daughter-cell nuclei were recorded. Only nuclei within a 150 μm² area centred at the dorsal midline and equator were used for analysis. The local curvature of the embryo within this region was minimal and thus not considered in our analysis. Each embryo was mounted such that the ML axis corresponds to the *x* axis, the AV axis corresponds to the *y* axis, and the depth axis corresponds to the *z* axis of the images (Fig. 1b). A line was drawn between the two daughter nuclei and the angle between this line and the AV axis was calculated. As the angle is in three dimensions, we broke it down into two planar projections: one in the epiblast plane (α) and one in the plane perpendicular to it (β). 0° represents divisions whose planar components are parallel to the AV axis, whereas 90° represents divisions whose planar components are orthogonal to the AV axis (Fig. 1b). All calculations were performed using custom routines written in MATLAB.

Analysis of cell elongation orientation

Cell shape was determined in interphase immediately before mitosis (approximately 12 min before anaphase). From each three-dimensional stack, the frame across the centre

of the nucleus was selected for analysis. Length to width ratios and cell elongation angles were measured using the best-fit ellipse function in NIH image.

Received 7 May; accepted 30 June 2004; doi:10.1038/nature02796.

Published online 14 July 2004.

- Ahringer, J. Control of cell polarity and mitotic spindle positioning in animal cells. *Curr. Opin. Cell Biol.* **15**, 73–81 (2003).
- Sausedo, R. A., Smith, J. L. & Schoenwolf, G. C. Role of nonrandomly oriented cell division in shaping and bending of the neural plate. *J. Comp. Neurol.* **381**, 473–488 (1997).
- Schoenwolf, G. C. & Alvarez, I. S. Roles of neuroepithelial cell rearrangement and division in shaping of the avian neural plate. *Development* **106**, 427–439 (1989).
- Wei, Y. & Mikawa, T. Formation of the avian primitive streak from spatially restricted blastoderm: evidence for polarized cell division in the elongating streak. *Development* **127**, 87–96 (2000).
- Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B. & Schilling, T. F. Stages of embryonic development of the zebrafish. *Dev. Dyn.* **203**, 253–310 (1995).
- Kimmel, C. B., Warga, R. M. & Kane, D. A. Cell cycles and clonal strings during formation of the zebrafish central nervous system. *Development* **120**, 265–276 (1994).
- Woo, K. & Fraser, S. E. Order and coherence in the fate map of the zebrafish nervous system. *Development* **121**, 2595–2609 (1995).
- Concha, M. L. & Adams, R. J. Oriented cell divisions and cellular morphogenesis in the zebrafish gastrula and neurula: a time-lapse analysis. *Development* **125**, 983–994 (1998).
- Hertwig, O. Ueber den Werth der ersten Furchungszellen für die Organbildung des Embryo. Experimentelle Studien am Froschund Tritonei. *Arch. Mikrosk. Anat.* **42**, 662–804 (1893).
- Black, S. D. & Vincent, J. P. The first cleavage plane and the embryonic axis are determined by separate mechanisms in *Xenopus laevis*. II. Experimental dissociation by lateral compression of the egg. *Dev. Biol.* **128**, 65–71 (1988).
- Wallingford, J. B. et al. Dishevelled controls cell polarity during *Xenopus* gastrulation. *Nature* **405**, 81–85 (2000).
- Tada, M. & Smith, J. C. Xwnt11 is a target of *Xenopus* Brachyury: regulation of gastrulation movements via Dishevelled, but not through the canonical Wnt pathway. *Development* **127**, 2227–2238 (2000).
- Jessen, J. R. et al. Zebrafish trilobite identifies new roles for Strabismus in gastrulation and neuronal movements. *Nature Cell Biol.* **4**, 610–615 (2002).
- Rothbacher, U. et al. Dishevelled phosphorylation, subcellular localization and multimerization regulate its role in early embryogenesis. *EMBO J.* **19**, 1010–1022 (2000).
- Hashimoto, H. et al. Zebrafish Dkk1 functions in forebrain specification and axial mesoderm formation. *Dev. Biol.* **217**, 138–152 (2000).
- Tago, K. et al. Inhibition of Wnt signaling by ICAT, a novel β-catenin-interacting protein. *Genes Dev.* **14**, 1741–1749 (2000).
- He, X., Semenov, M., Kamai, K. & Zeng, X. LDL receptor-related proteins 5 and 6 in Wnt/β-catenin signalling: Arrows point the way. *Development* **131**, 1663–1677 (2004).
- Heisenberg, C. P. et al. Silberblick/Wnt11 mediates convergent extension movements during zebrafish gastrulation. *Nature* **405**, 76–81 (2000).
- Kilian, B. et al. The role of Ppt/Wnt5 in regulating cell shape and movement during zebrafish gastrulation. *Mech. Dev.* **120**, 467–476 (2003).
- Moon, R. T. et al. Xwnt-5A: a maternal Wnt that affects morphogenetic movements after overexpression in embryos of *Xenopus laevis*. *Development* **119**, 97–111 (1993).
- Park, M. & Moon, R. T. The planar cell-polarity gene stbm regulates cell behaviour and cell fate in vertebrate embryos. *Nature Cell Biol.* **4**, 20–25 (2002).
- Glickman, N. S., Kimmel, C. B., Jones, M. A. & Adams, R. J. Shaping the zebrafish notochord. *Development* **130**, 873–887 (2003).
- Elul, T. & Keller, R. Monopolar protrusive activity: a new morphogenic cell behavior in the neural plate dependent on vertical interactions with the mesoderm in *Xenopus*. *Dev. Biol.* **224**, 3–19 (2000).
- Sausedo, R. A. & Schoenwolf, G. C. Quantitative analyses of cell behaviors underlying notochord formation and extension in mouse embryos. *Anat. Rec.* **239**, 103–112 (1994).
- Schlesinger, A., Shelton, C. A., Maloof, J. N., Meneghini, M. & Bowerman, B. Wnt pathway components orient a mitotic spindle in the early *Caenorhabditis elegans* embryo without requiring gene transcription in the responding cell. *Genes Dev.* **13**, 2028–2038 (1999).
- Gho, M. & Schweisguth, F. Frizzled signalling controls orientation of asymmetric sense organ precursor cell divisions in *Drosophila*. *Nature* **393**, 178–181 (1998).
- Chalmers, A. D., Strauss, B. & Papalopulu, N. Oriented cell divisions asymmetrically segregate aPKC and generate cell fate diversity in the early *Xenopus* embryo. *Development* **130**, 2657–2668 (2003).
- Das, T., Payer, B., Cayouette, M. & Harris, W. A. In vivo time-lapse imaging of cell divisions during neurogenesis in the developing zebrafish retina. *Neuron* **37**, 597–609 (2003).
- Koster, R. W. & Fraser, S. E. Tracing transgene expression in living zebrafish embryos. *Dev. Biol.* **233**, 329–346 (2001).
- Campbell, R. E. et al. A monomeric red fluorescent protein. *Proc. Natl Acad. Sci. USA* **99**, 7877–7882 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank M. Hibi (Zdkk1), M. Park (Stbm MO), J. Wallingford (Xdsh-D2), R. Köster (H2B-GFP), S. Megason (H2B-RFP1 and membrane RFP1), M. Tada (Xdsh-ΔPDZ and Xdsh-DEP +) and H. McBride (ICAT) for providing the plasmids and morpholino, and C.-P. Heisenberg for the *silberblick/wnt11* mutants. We also thank S. Bhattacharyya, M. Bronner-Fraser, M. Garcia-Castro, D. Koos, R. Köster, B. Link, S. Megason and J. Wallingford for discussion and critical reading of the manuscript.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.E.F. (sefraser@caltech.edu).

De-ubiquitination and ubiquitin ligase domains of A20 downregulate NF- κ B signalling

Ingrid E. Wertz^{1,4}, Karen M. O'Rourke¹, Honglin Zhou¹, Michael Eby², L. Aravind⁵, Somasekar Seshagiri², Ping Wu³, Christian Wiesmann³, Rohan Baker⁶, David L. Boone⁷, Averil Ma⁷, Eugene V. Koonin⁵ & Vishva M. Dixit¹

¹Department of Molecular Oncology, ²Department of Molecular Biology and

³Department of Protein Engineering, Genentech Inc., South San Francisco, California 94080, USA

⁴Department of Biological Chemistry, School of Medicine, University of California, Davis, Davis, California 95616, USA

⁵Computational Biology Branch, NCBI, NLM, NIH, Bethesda, Maryland 20894, USA

⁶Division of Molecular Bioscience, John Curtin School of Medical Research, Australian National University, Canberra, Australian Capital Territory 2601, Australia

⁷Department of Medicine, University of California, San Francisco, San Francisco, California 94143, USA

NF- κ B transcription factors mediate the effects of pro-inflammatory cytokines such as tumour necrosis factor- α and interleukin-1 β ¹. Failure to downregulate NF- κ B transcriptional activity results in chronic inflammation and cell death, as observed in A20-deficient mice². A20 is a potent inhibitor of NF- κ B signalling, but its mechanism of action is unknown². Here we show that A20 downregulates NF- κ B signalling through the cooperative activity of its two ubiquitin-editing domains. The amino-terminal domain of A20, which is a de-ubiquitinating (DUB) enzyme of the OTU (ovarian tumour) family³, removes lysine-63 (K63)-linked ubiquitin chains from receptor interacting protein (RIP), an essential mediator of the proximal TNF receptor 1 (TNFR1) signalling complex^{4,5}. The carboxy-terminal domain of A20, composed of seven C₂/C₂ zinc fingers⁶, then functions as a ubiquitin ligase by polyubiquitinating RIP with K48-linked ubiquitin chains, thereby targeting RIP for proteasomal degradation. Here we define a novel ubiquitin ligase domain and identify two sequential mechanisms by which A20 downregulates NF- κ B signalling. We also provide an example of a protein containing separate ubiquitin ligase and DUB domains, both of which participate in mediating a distinct regulatory effect.

To explore the mechanism by which A20 downregulates NF- κ B signalling, we analysed proteins from multiple species that contained A20-type zinc fingers (ZnFs) (Supplementary Fig. S1) or the OTU domain. A20-like ZnFs and OTU domains were often associated with domains involved in the ubiquitin system^{3,7} (Supplementary Fig. S2), raising the possibility that A20 participates in ubiquitin signalling pathways.

We therefore investigated whether A20 was associated with ubiquitin ligase activity. Ubiquitin ligases catalyse the final step of substrate protein ubiquitination and are characterized by two structurally unrelated ubiquitin transferring domains, the RING finger and the HECT domain^{8,9}. FLAG-tagged A20 was immunoprecipitated from transfected HEK293T cells and FLAG peptide-eluted proteins were added to *in vitro* ubiquitination assays. FLAG-A20 and co-eluting proteins catalysed potent polyubiquitination with specific ubiquitin-conjugating (E2) enzymes only when all reaction components were included (Fig. 1a; left panels in Fig. 1b). To determine whether a co-eluting ubiquitin ligase was responsible for the A20-associated catalytic activity, recombinant A20 was prepared from *Escherichia coli*, that lack the ubiquitin system. Remarkably, despite the absence of a RING finger or HECT domain, recombinant A20 catalysed *in vitro* polyubiquitination

(Fig. 1b, right panels), which was mediated by the ZnF domain (Fig. 1c). Next, A20 truncation mutants were analysed to map the ZnF(s) responsible for catalysing polyubiquitination. Collectively, these analyses indicated that ZnFs 3 and/or 4 are required for catalysis (Supplementary Figs S3 and S4). To test this idea, full-length A20 with point mutations of conserved cysteines within ZnF3 (C521A, C524A) or ZnF4 (C624A, C627A) was purified from HEK293T cells (Supplementary Fig. S5) or from *E. coli* (Fig. 1d). The A20 ZnF4 mutant, but not the ZnF3 mutant, markedly attenuated A20 ubiquitin ligase activity from both sources, implicating ZnF4 as a critical moiety for ubiquitin transfer (Supplementary Fig. S5; Fig. 1d). Having established that A20 contains a novel ubiquitin ligase domain, our next aim was to identify A20 substrates. We first investigated whether A20 catalysed the formation of K48-linked polyubiquitin chains, which target substrates for proteasomal degradation, or whether A20 catalysed K63-linked polyubiquitin chains, which regulate substrate activity but do not promote effective degradation^{8,9}. The use of ubiquitin mutants (Fig. 2a) with only K48 (K48-only Ub) or K63 (K63-only Ub) available for polymerization demonstrated that A20 catalyses K48-linked polyubiquitination (Fig. 2b), suggesting that A20 promotes substrate degradation.

As cells from A20^{-/-} mice failed to downregulate tumour necrosis factor- α (TNF- α), but not interleukin-1 β (IL-1 β)-induced NF- κ B signalling², we speculated that A20 substrates are positive regulators of NF- κ B signalling unique to TNF- α -activated pathways. TNF receptor 1-associated protein (TRADD), RIP, tumour necrosis factor receptor associated factor 2 (TRAF2) and TRAF5 are candidates that fit these criteria; however, all these proteins (except for RIP) are involved in both NF- κ B and c-jun N-terminal kinase (JNK) signalling¹⁰⁻¹³. RIP does not participate in JNK signalling^{4,5}, yet downregulation of JNK signalling was only protracted in A20^{-/-} cells, whereas NF- κ B activity remained constitutively elevated². These data therefore pointed to RIP as a potential A20 substrate. Indeed, previous studies showed that RIP and A20 are recruited to the activated TNFR1 receptor complex^{14,15}, that recruited RIP is potentially ubiquitinated upon TNFR1 recruitment¹⁶ and that levels of RIP associated with TNFR1 diminish over time¹⁵⁻¹⁷. To assess whether TNFR1-associated RIP is targeted for proteasomal degradation, HeLaS3 cells were treated with TNF- α in the presence or absence of the proteasome inhibitor MG-132, and cytosolic and TNFR1-associated RIP were analysed by immunoblotting. MG-132 treatment stabilized TNFR1-associated RIP, but cytosolic RIP levels did not fluctuate, consistent with a previous report¹⁶ (Supplementary Fig. S6). Treatment of TNFR1 immunoprecipitates with the non-specific DUB enzyme Usp2 confirmed that TNFR1-associated RIP is ubiquitinated (data not shown). Taken together, these results suggest that a ubiquitin ligase targets TNFR1-associated RIP for proteasomal degradation. To evaluate whether A20 directly ubiquitinates RIP, recombinant A20 expressed in *E. coli* and recombinant RIP expressed in insect cells were purified for *in vitro* ubiquitination assays. A20 directly ubiquitinated RIP, thereby establishing RIP as an A20 substrate (Fig. 2c). Co-transfected A20 also promoted RIP degradation (Fig. 2d). Notably, point mutations within A20 ZnF4, which attenuate A20-mediated ubiquitination (Fig. 1d), but not point mutations within A20 ZnF3, which do not affect A20-mediated ubiquitination (Fig. 1d), inhibited A20-mediated RIP degradation (Fig. 2d). Because wild-type, ZnF3- and ZnF4-mutant A20 all bound endogenous RIP equally well (Fig. 2e), these results suggested that A20-mediated RIP ubiquitination and degradation were functionally linked and dependent on ZnF4. To assess the physiological significance of A20 ZnF4 in regulating NF- κ B signalling, A20^{-/-} murine embryonic fibroblasts (MEFs) were complemented with wild-type or ZnF4 mutant A20. Although wild-type A20 effectively attenuated TNF- α -induced NF- κ B signalling, the A20 ZnF4 mutant did not (Fig. 2f). Collectively, these results imply that A20 downregulates TNF- α -induced NF- κ B signalling by catalysing RIP ubiquitination,

thereby promoting its subsequent degradation. Having identified a function for the C-terminal A20 ZnFs, we next focused on the N-terminal OTU domain of A20. Four independent reports recently confirmed the predicted DUB activity of the OTU domain³ and the indispensability of the catalytic cysteine for this reaction^{18–21}. However, no biological substrates for these DUB enzymes were identified. To test whether A20 is a DUB, K48- and K63-linked ubiquitin chains were synthesized as substrates for *in vitro* assays. Recombinant A20 effectively depolymerized both types of ubiquitin chains, thereby establishing A20 as a DUB (Fig. 3a).

As RIP is a substrate for the A20 ZnF domain (Fig. 2c, d), we

investigated whether the A20 DUB domain also modifies RIP. We first assessed the effect of A20 on RIP ubiquitination by co-transfecting Myc-RIP and HA-ubiquitin with increasing doses of wild-type- or OTU mutant-A20. Myc-RIP was ubiquitinated on overexpression (Fig. 3b). Intriguingly, wild-type A20 markedly decreased Myc-RIP ubiquitination, whereas the OTU-mutant A20 did not (Fig. 3b), even though wild-type and OTU-mutant A20 bound endogenous RIP equally well (Fig. 3c). Furthermore, recombinant A20 removed ubiquitin chains from RIP *in vitro*, thereby establishing RIP as a direct substrate of the A20 OTU domain (Fig. 3d). Given the importance of K63-linked ubiquitination in

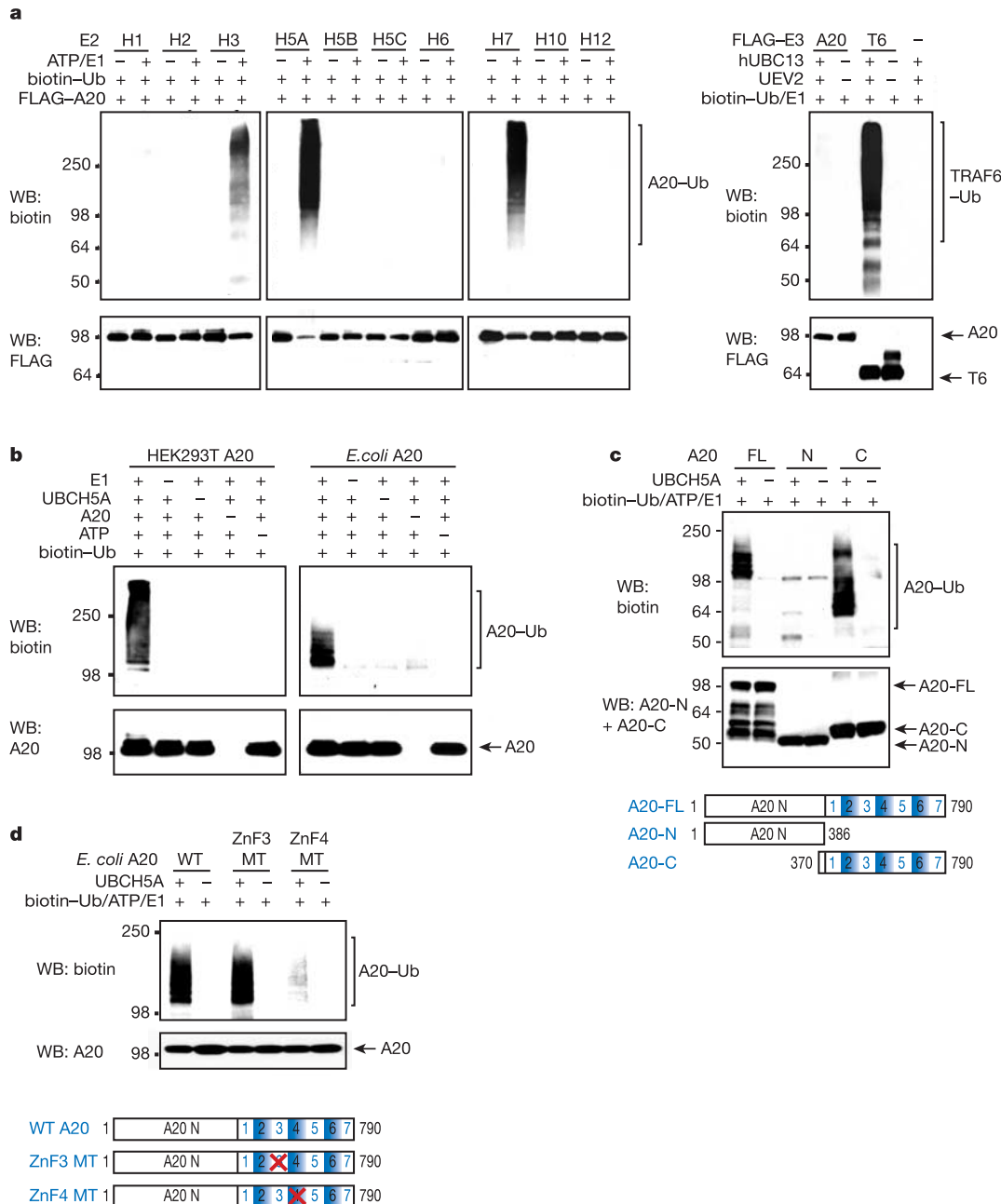


Figure 1 A20 catalyses *in vitro* ubiquitination. **a**, *In vitro* ubiquitination assays with FLAG-eluted A20 from HEK293T cells and a panel of ubiquitin conjugating (E2) enzymes. Ubiquitinated proteins are detected by anti-biotin western blot (WB). T6: TRAF6, a positive control ubiquitin ligase (E3) for hUBC13/UEV2-mediated polyubiquitination. Ub: ubiquitin. **b**, A20 FLAG-eluted from HEK293T cells (left panels) or purified from *E. coli* (right panels)

has intrinsic ubiquitin ligase activity *in vitro* when all reaction components are supplied. **c**, A20 ZnFs catalyse *in vitro* polyubiquitination. A20 proteins were purified from *E. coli*. FL: full length, N: N-terminal, C: C-terminal. **d**, Point mutations of conserved cysteines in A20 ZnF4 (ZnF4MT), but not ZnF3 (ZnF3MT), attenuate *in vitro* ubiquitination catalysed by A20 purified from *E. coli*. All markers on the left hand side are in kDa.

activating NF- κ B signalling^{22–26}, we hypothesized that ubiquitination of transfected RIP was K63-linked, and that the A20 OTU domain removed these K63-linked ubiquitin chains. This scenario is not without precedence, because the tumour suppressor cylindromatosis (CYLD) also regulates NF- κ B signalling by de-ubiquitinating K63-linked chains on TRAF2, TRAF6 and NEMO²⁵.

We therefore investigated ubiquitin ligases that might promote K63-linked RIP ubiquitination. TRAF2, which catalyses K63-linked ubiquitination with the heterodimeric E2 hUBC13/UEV (ref. 24), is recruited to the activated TNFR1 signalling complex²⁷, and activates NF- κ B signalling¹³ in a hUBC13-dependent manner²². Indeed, TRAF2 expression promoted ubiquitination of endogenous RIP with K63-linked, but not K48-linked chains, and A20 effectively removed TRAF2-mediated RIP ubiquitination (Fig. 3e). The related ubiquitin ligase TRAF6, which participates in toll-like receptor (TLR) and IL-1 β -induced NF- κ B signalling, but not in TNF- α -activated NF- κ B signalling¹, did not promote RIP ubiquitination, suggesting that TRAF2-induced RIP ubiquitination was specific (Fig. 3e). Additionally, attenuation of endogenous TRAF2 or hUBC13 expression with small interfering RNA oligonucleotides (siRNA) reduced ubiquitination of transfected RIP (Fig. 3f). Thus TRAF2 promotes K63-linked RIP ubiquitination, whereas A20 catalyses K48-linked RIP ubiquitination.

To determine the physiological significance of the A20 OTU domain in regulating NF- κ B signalling, we complemented A20^{-/-}

MEFs with wild-type A20 or the A20 OTU catalytic cysteine mutant (C103A). A20^{-/-} MEFs complemented with the A20 OTU mutant were less effective in down-regulating TNF- α -induced NF- κ B activity than wild-type A20 (Fig. 3g), consistent with a recent study¹⁸. Our results therefore suggested, perhaps counter-intuitively, that downregulation of NF- κ B required two opposing reactions catalysed by A20: RIP ubiquitination and de-ubiquitination.

Given this apparent paradox, we hypothesized that one A20 ubiquitin-editing domain might regulate the activity of the other. We therefore tested whether A20-mediated RIP de-ubiquitination regulated A20-induced RIP degradation. Whereas wild-type A20 effectively promoted RIP degradation, an A20 mutant lacking a functional DUB domain as a result of deletion (Fig. 4a) or mutation (Fig. 4b) promoted less RIP degradation. These results therefore suggested that A20-mediated de-ubiquitination of K63-linked ubiquitin chains on RIP (Fig. 3e) is a prerequisite for A20-mediated RIP degradation through K48-linked ubiquitin chains (Fig. 2b, d).

To test this hypothesis, we co-transfected RIP with increasing doses of wild-type, OTU mutant or ZnF4 mutant A20 and ubiquitin mutants in which only K48 or K63 are available for polymerization (Fig. 2a). Wild-type A20 disassembled K63-linked ubiquitin chains on RIP and promoted K48-linked RIP ubiquitination (Fig. 4c). The A20 OTU mutant markedly attenuated de-ubiquitination of K63-linked ubiquitin chains on RIP, implicating the OTU domain

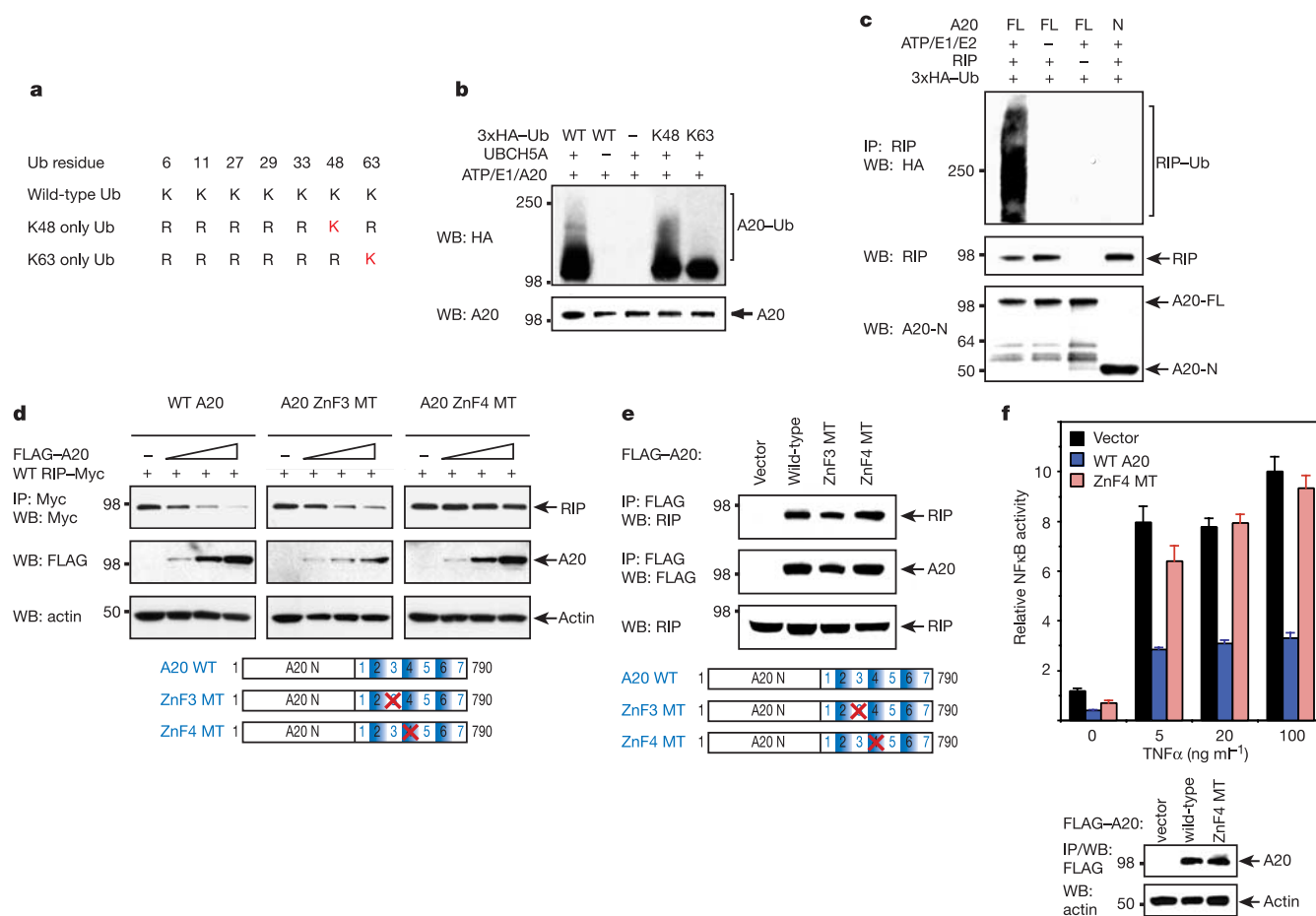


Figure 2 A20 ubiquitinates and destabilizes RIP. **a**, Schematic diagram of ubiquitin mutants. **b**, A20 purified from *E. coli* autoubiquitinates with K48-only, but not K63-only, polyubiquitin chains *in vitro*. **c**, A20 purified from *E. coli* directly ubiquitinates RIP purified from insect cell lysates *in vitro*. **d**, Co-transfection of A20 promotes RIP degradation in HEK293T cells. A20 ZnF4MT, but not ZnF3MT, attenuates RIP degradation.

e, Endogenous RIP binding to transfected FLAG-A20 variants in HEK293T cells. **f**, Complementation of A20^{-/-} MEFs with wild-type, but not ZnF4MT, A20 effectively downregulates NF- κ B signalling (upper graph). Average values are of four independent transfections ($\pm$ s.d.). Relative expression levels of transfected constructs (lower western blots). All markers on the left hand side are in kDa.

as the essential moiety for DUB activity (Fig. 4c). Mutation of the A20 ZnF4 domain markedly diminished the ability of A20 to ubiquitinate RIP with K48-linked ubiquitin chains, thereby implicating ZnF4 as the critical ubiquitin ligase domain (Fig. 4c). The ZnF4 domain mutant, which has a functional OTU domain, effectively removed K63-linked ubiquitin chains from RIP; in contrast, the OTU domain mutant, which has a functional ZnF4 domain, was defective in adding K48-linked ubiquitin chains on RIP (Fig. 4c). These results support the hypothesis that RIP de-ubiquitination is a prerequisite for A20-mediated RIP degradation.

Furthermore, these results indicate that A20 has two distinct catalytic domains, both of which cooperate to downregulate NF- κ B signalling: the N-terminal OTU domain removes K63-linked ubiquitin chains from active RIP, which then permits the C-terminal ZnF region to target RIP for proteasomal degradation through K48-linked polyubiquitination.

If de-ubiquitination of K63-linked ubiquitin chains on RIP precedes polyubiquitination with K48-linked ubiquitin chains, it is predicted that, even in the absence of A20, RIP recruited to the activated TNFR1 complex will be ubiquitinated by TRAF2 with

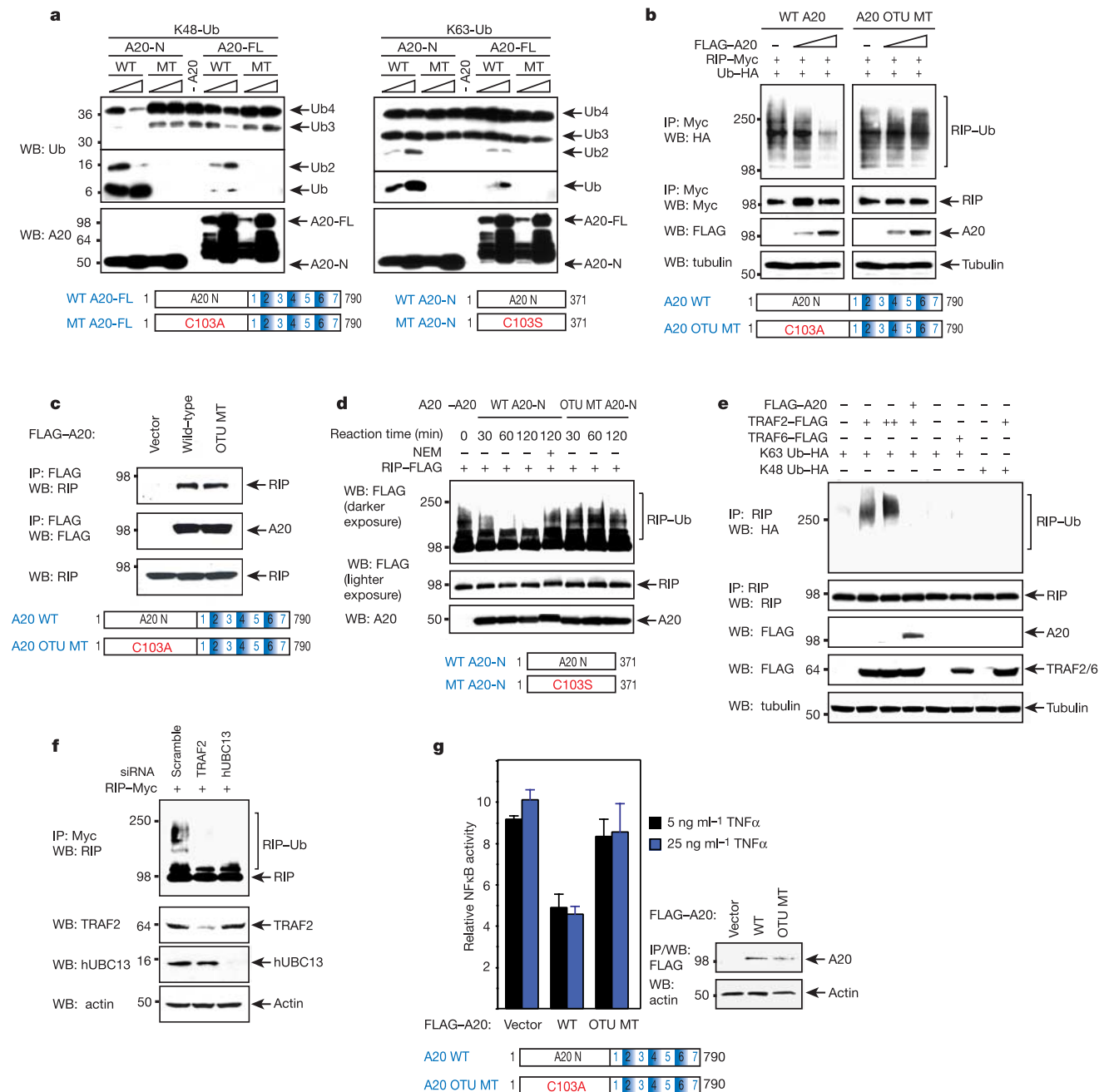


Figure 3 A20 de-ubiquitinates RIP. **a**, A20 purified from *E. coli* depolymerizes free polyubiquitin chains *in vitro*. **b**, Co-transfection of wild-type A20 de-ubiquitinates RIP in HEK293T cells. **c**, Transfected wild-type and OTU mutant FLAG-A20 bind endogenous RIP equally well in HEK293T cells. **d**, A20 purified from *E. coli* de-ubiquitinates FLAG-RIP purified from HEK293T cells *in vitro*. NEM: N-ethylmaleimide, a DUB inhibitor. **e**, Transfection of TRAF2 specifically promotes K63-linked ubiquitination of endogenous

RIP in HEK293T cells. **f**, Ubiquitination of transfected RIP is reduced by attenuation of endogenous TRAF2 or hUBC13 expression with small interfering RNA oligonucleotides (siRNA) in HEK293T cells. **g**, The A20 OTU domain is required for A20-mediated NF- κ B downregulation in complemented *A20*^{-/-} MEFs (left graph). Average values are of three independent transfections ($\pm$ s.d.). Relative expression levels of transfected constructs (right Western blots). All markers on the left hand side are in kDa.

K63-linked chains (Fig. 3e, f). However, in the absence of A20, RIP will neither be de-ubiquitinated nor targeted for proteasomal degradation. Indeed, RIP recruited to activated TNFR1 remained hyperubiquitinated and was stabilized in $A20^{-/-}$ MEFs (Fig. 4d).

To investigate the respective roles of each A20 domain in regulating endogenous RIP ubiquitination and degradation at TNFR1, we analysed TNFR1-associated RIP in $A20^{-/-}$ MEFs complemented with wild-type A20, OTU mutant A20 or ZnF4 mutant A20. Whereas RIP levels in vector-complemented $A20^{-/-}$ MEFs remained stable, complementation with wild-type A20 promoted degradation of TNFR1-associated RIP over time (Fig. 4e). In A20 OTU mutant-complemented $A20^{-/-}$ MEFs, RIP recruited to

TNFR1 remained hyperubiquitinated and was stabilized (Fig. 4e). Complementation with the A20 ZnF4 mutant resulted in some RIP de-ubiquitination, probably mediated by the intact A20 OTU domain, yet blocked degradation of TNFR1-associated RIP (Fig. 4e), consistent with the inability of the A20 ZnF4 domain to catalyse effective K48-linked RIP ubiquitination (Fig. 4c) and degradation (Fig. 2d). These results confirm the essential functions of the A20 OTU and ZnF4 domains in regulating TNFR1-associated RIP degradation and TNF- α -induced NF- κ B activity (Figs 2f and 3g).

Herein we identify two mechanisms by which A20 negatively regulates NF- κ B signalling: (1) the disassembly of K63-linked

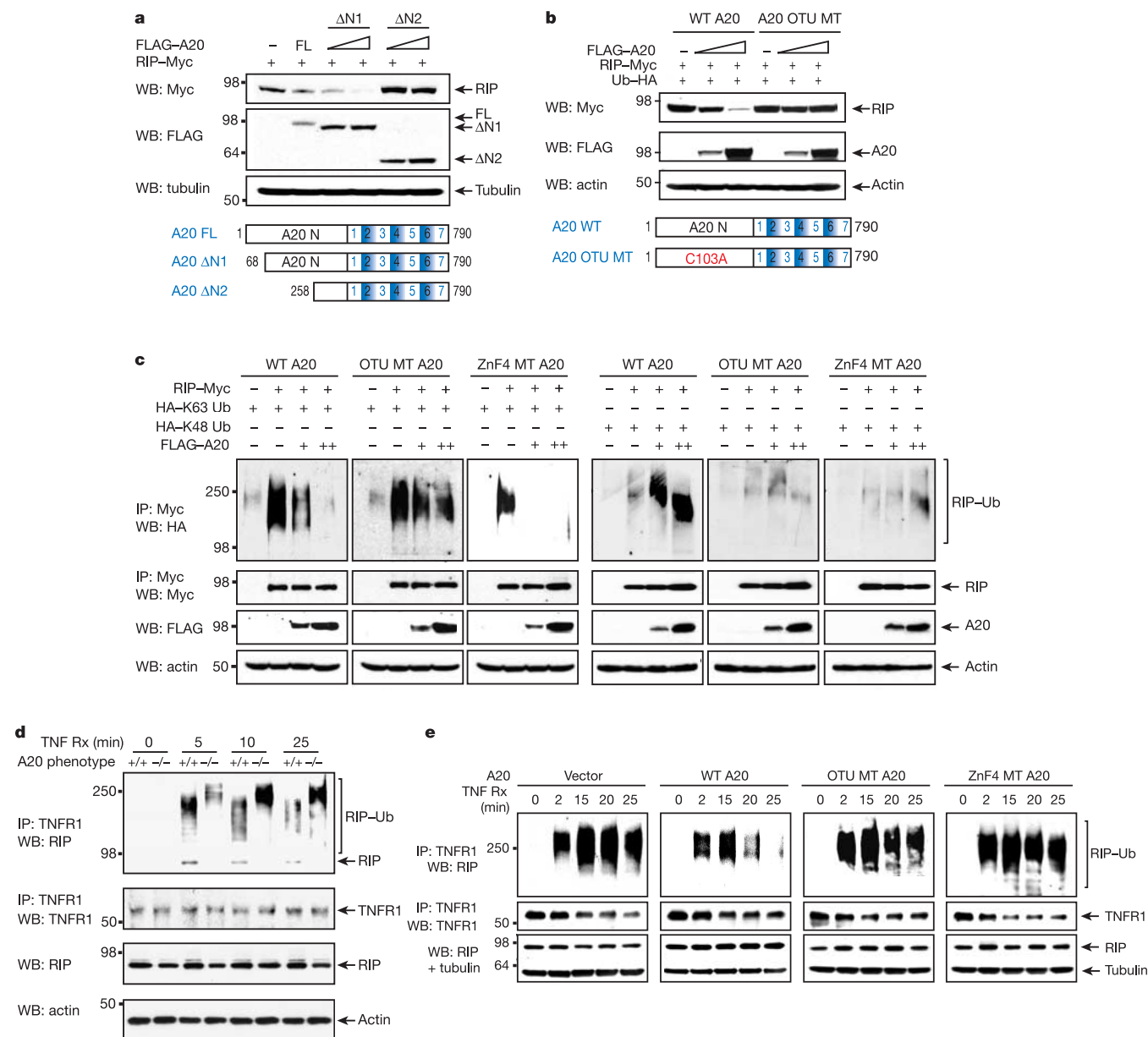


Figure 4 A20-mediated RIP de-ubiquitination is a prerequisite for A20-induced degradation. **a**, Deletion of the A20 OTU domain blocks destabilization of co-transfected RIP in HEK293T cells. **b**, Mutation of the A20 OTU domain blocks destabilization of co-transfected RIP in HEK293T cells. **c**, Functional interplay of the A20 OTU and ZnF4 domains in co-transfected HEK293T cells. De-ubiquitination of K63-linked chains on RIP by the A20 OTU domain is a prerequisite for RIP polyubiquitination with K48-linked chains

by A20 ZnF4. **d**, Endogenous RIP recruited to TNFR1 is hyperubiquitinated and stabilized in $A20^{-/-}$ MEFs. **e**, Both the A20 OTU domain and the A20 ZnF4 domains regulate the stability of TNFR1-associated endogenous RIP in complemented $A20^{-/-}$ MEFs. In panels **d** and **e**, 1/200 of the total cell lysate was used for straight WB analysis. All markers on the left hand side are in kDa.

ubiquitin chains bound to the essential adaptor RIP, and (2) the targeting of RIP for degradation through the ligation of K48-linked ubiquitin chains (Supplementary Fig. S7). It is also possible that A20 regulates other substrates in NF- κ B signalling pathways in a similar manner. Previous studies have shown that K63-linked TRAF6 autoubiquitination activates the TAK1 kinase by an unknown mechanism²³ and that K63-linked TRAF2 autoubiquitination activates GCKR and SAPK kinases²⁴. Because the RIP kinase domain is dispensable for NF- κ B activation^{4,14,28}, the exact function of K63-linked ubiquitination of RIP is unclear. However, studies with RIP-null cells clearly implicate RIP as an essential mediator of TNF- α -induced signalling^{4,5,28}, possibly through recruiting, oligomerizing and thereby promoting activation of the I κ B kinase (IKK) complex^{15,29,30}. Because K63-linked ubiquitin chains are critical for mediating NF- κ B activation^{22–26}, it is plausible that K63-linked ubiquitinated RIP serves as a nexus for the assembly of the activated TNFR1 and IKK signalling complexes, possibly through interactions with the ubiquitin-related domains in the I κ B- α and I κ B- β proteins⁷.

The A20-type ZnF represents a novel class of ubiquitin ligases. A20 ZnF4 is characterized by an acidic residue in the third position and a large residue in the eighth position distal to the fourth conserved cysteine (Supplementary Fig. S1). This acidic residue and the polar helix on which it is located could form a critical interaction surface required for ubiquitin transfer. A20-type ZnFs are found in all eukaryotes and form fusions with a variety of domains involved in ubiquitin signalling pathways^{3,7} (Supplementary Fig. S2). Thus, the A20 ZnF domain probably represents an ancient, conserved ubiquitin transferring module that functions in diverse biological contexts. Our characterization of A20 as a dual-function ubiquitin-editing enzyme, with distinct peptidase and ligase domains, suggests that such enzymes might be a widespread feature of the ubiquitin signalling system. □

Methods

Details regarding the plasmids and reagents, the cell culture and transfection conditions, the immunoprecipitation and immunoblotting protocols, the protein purifications and FLAG elutions, and the protocols briefly described below can be found in Supplementary Information.

In vivo ubiquitination and de-ubiquitination assays

HEK293T cells were transfected as indicated and pre-treated with 25 μ M MG-132. 10 mM N-ethylmaleimide and 25 μ M MG-132 were added to the lysis buffer, and lysates were cleared by centrifugation. Proteins were dissociated by heating at 90 °C in 1% SDS (v/v) and samples were diluted 1:10. Myc-RIP was immunoprecipitated with anti-Myc agarose (Covance), and endogenous RIP was immunoprecipitated with a cocktail of RIP antibodies and Protein A/G Plus beads (Santa Cruz). Immunoprecipitates were washed and prepared for immunoblot analysis as indicated.

In vitro ubiquitination assays

Autoubiquitination assays were performed in 50 μ l reaction volumes with the following components as indicated: 2 μ g N-terminal biotinylated ubiquitin (Boston Biochem) or 3 $\times$ HA-tagged ubiquitin, 0.2 μ g E1 (Calbiochem), 1.0 μ g E2 (Boston Biochem) or 0.5 μ g each of hUBC13 and MMS2, 15 μ l (up to 1 μ g) E3 and 5 μ l 10 $\times$ reaction buffer (300 mM HEPES at pH 7.2, 20 mM ATP, 50 mM MgCl₂ and 2 mM dithiothreitol). Reactions were incubated at 30 °C for 1 h and prepared for immunoblot analysis as indicated. For *in vitro* RIP ubiquitination by A20, reactions were performed as for autoubiquitination assays except 1 μ g recombinant RIP was included and 5 μ g 3 $\times$ HA-tagged ubiquitin was used instead of biotinylated ubiquitin. Reactions were performed in quadruplicate, recombined, and 50 μ l was reserved for immunoblotting. Proteins in the remaining 150 μ l were dissociated by heating at 90 °C in 1% SDS (v/v) and were diluted 1:10. RIP was immunoprecipitated by incubation with a cocktail of RIP antibodies and Protein A/G Plus beads (Santa Cruz) and immunoprecipitates were washed and prepared for immunoblot analysis as indicated.

In vitro de-ubiquitination assays

FLAG-RIP substrate, polyubiquitin chains and A20 proteins were purified as described in the Supplementary Information. FLAG-RIP immunoprecipitates were washed twice with dissociation buffer, once with PBS and twice with buffer containing 50 mM HEPES at pH 8.0 and 0.01% Brij-35. Enzymes (up to 1 μ g) and substrates (up to 1 μ M) were combined in a buffer containing 50 mM HEPES at pH 8.0, 0.01% Brij-35 and 3 mM DTT and incubated at 37 °C for 150 min (free polyubiquitin chains) or for the indicated time (FLAG-RIP substrate). Samples were prepared for immunoblot analysis as indicated.

RNA interference

TRAF2 siRNA duplexes (5'-CGACAUGAACAUCCGCAAGC-3') with 3' dTdT overhangs were synthesized at Genentech. Scramble II siRNA duplexes were purchased from Dharmacon and hUBC13 siRNA duplexes have been described²⁶. HEK293T cells were transfected with siRNAs three times at 24-h intervals. Before the third transfection, cells were split and co-transfected with the indicated siRNAs and mammalian expression constructs. Cells were collected 30 h thereafter for analysis.

Received 22 April; accepted 29 June 2004; doi:10.1038/nature02794.

Published online 18 July 2004.

- Dixit, V. & Mak, T. W. NF- κ B signalling. Many roads lead to Madrid. *Cell* **111**, 615–619 (2002).
- Lee, E. G. *et al.* Failure to regulate TNF-induced NF- κ B and cell death responses in A20-deficient mice. *Science* **289**, 2350–2354 (2000).
- Makarova, K. S., Aravind, L. & Koonin, E. V. A novel superfamily of predicted cysteine proteases from eukaryotes, viruses and *Chlamydia pneumoniae*. *Trends Biochem. Sci.* **25**, 50–52 (2000).
- Ting, A. T., Pimentel-Muinos, F. X. & Seed, B. RIP mediates tumor necrosis factor receptor 1 activation of NF- κ B but not Fas/APO-1-initiated apoptosis. *EMBO J.* **15**, 6189–6196 (1996).
- Kelliher, M. A. *et al.* The death domain kinase RIP mediates the TNF-induced NF- κ B signal. *Immunity* **8**, 297–303 (1998).
- Opipari, A. W. Jr, Boguski, M. S. & Dixit, V. M. The A20 cDNA induced by tumor necrosis factor alpha encodes a novel type of zinc finger protein. *J. Biol. Chem.* **265**, 14705–14708 (1990).
- Aravind, L., Dixit, V. M. & Koonin, E. V. Apoptotic molecular machinery: vastly increased complexity in vertebrates revealed by genome comparisons. *Science* **291**, 1279–1284 (2001).
- Pickart, C. M. Mechanisms underlying ubiquitination. *Annu. Rev. Biochem.* **70**, 503–533 (2001).
- Weissman, A. M. Themes and variations on ubiquitination. *Nature Rev. Mol. Cell Biol.* **2**, 169–178 (2001).
- Yeh, W. C. *et al.* Early lethality, functional NF- κ B activation, and increased sensitivity to TNF-induced cell death in TRAF2-deficient mice. *Immunity* **7**, 715–725 (1997).
- Lee, S. Y. *et al.* TRAF2 is essential for JNK but not NF- κ B activation and regulates lymphocyte proliferation and survival. *Immunity* **7**, 703–713 (1997).
- Kieser, A., Kaiser, C. & Hammerschmidt, W. LMP1 signal transduction differs substantially from TNF receptor 1 signaling in the molecular functions of TRADD and TRAF2. *EMBO J.* **18**, 2511–2521 (1999).
- Tada, K. *et al.* Critical roles of TRAF2 and TRAF5 in tumor necrosis factor-induced NF- κ B activation and protection from cell death. *J. Biol. Chem.* **276**, 36530–36534 (2001).
- Hsu, H., Huang, J., Shu, H. B., Baichwal, V. & Goeddel, D. V. TNF-dependent recruitment of the protein kinase RIP to the TNF receptor-1 signaling complex. *Immunity* **4**, 387–396 (1996).
- Zhang, S. Q., Kovalenko, A., Cantarella, G. & Wallach, D. Recruitment of the IKK signalosome to the p55 TNF receptor: RIP and A20 bind to NEMO (IKK γ) upon receptor stimulation. *Immunity* **12**, 301–311 (2000).
- Legler, D. F., Micheau, O., Doucey, M. A., Tschopp, J. & Bron, C. Recruitment of TNF receptor 1 to lipid rafts is essential for TNF α -mediated NF- κ B activation. *Immunity* **18**, 655–664 (2003).
- Micheau, O. & Tschopp, J. Induction of TNF receptor I-mediated apoptosis via two sequential signaling complexes. *Cell* **114**, 181–190 (2003).
- Balakirev, M. Y., Tcherniuk, S. O., Jaquinod, M. & Chroboczek, J. Otubains: a new family of cysteine proteases in the ubiquitin pathway. *EMBO Rep.* **4**, 517–522 (2003).
- Borodovsky, A. *et al.* Chemistry-based functional proteomics reveals novel members of the deubiquitinating enzyme family. *Chem. Biol.* **9**, 1149–1159 (2002).
- Evans, P. C. *et al.* A novel type of deubiquitinating enzyme. *J. Biol. Chem.* **278**, 23180–23186 (2003).
- Evans, P. C. *et al.* Zinc-finger protein A20, a regulator of inflammation and cell survival, has de-ubiquitinating activity. *Biochem. J.* **378**, 727–734 (2004).
- Deng, L. *et al.* Activation of the I κ B kinase complex by TRAF6 requires a dimeric ubiquitin-conjugating enzyme complex and a unique polyubiquitin chain. *Cell* **103**, 351–361 (2000).
- Wang, C. *et al.* TAK1 is a ubiquitin-dependent kinase of MKK and IKK. *Nature* **412**, 346–351 (2001).
- Shi, C. S. & Kehrl, J. H. Tumor necrosis factor (TNF)-induced germinal center kinase-related (GCKR) and stress-activated protein kinase (SAPK) activation depends upon the E2/E3 complex Ubc13-Uev1A/TNF receptor-associated factor 2 (TRAF2). *J. Biol. Chem.* **278**, 15429–15434 (2003).
- Wilkinson, K. D. Signal transduction: aspirin, ubiquitin and cancer. *Nature* **424**, 738–739 (2003).
- Zhou, H. *et al.* Bcl10 activates the NF- κ B pathway through ubiquitination of NEMO. *Nature* **427**, 167–171 (2004).
- Hsu, H., Shu, H. B., Pan, M. G. & Goeddel, D. V. TRADD-TRAF2 and TRADD-FADD interactions define two distinct TNF receptor 1 signal transduction pathways. *Cell* **84**, 299–308 (1996).
- Devin, A. *et al.* The distinct roles of TRAF2 and RIP in IKK activation by TNF-R1: TRAF2 recruits IKK to TNF-R1 while RIP mediates IKK activation. *Immunity* **12**, 419–429 (2000).
- Poyet, J. L. *et al.* Activation of the I κ B kinases by RIP via IKK γ /NEMO-mediated oligomerization. *J. Biol. Chem.* **275**, 37966–37977 (2000).
- Devin, A. *et al.* The alpha and beta subunits of I κ B kinase (IKK) mediate TRAF2-dependent IKK recruitment to tumor necrosis factor (TNF) receptor 1 in response to TNF. *Mol. Cell. Biol.* **21**, 3986–3994 (2001).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements The authors would like to thank R. Deshaies, T. Mayor, R. Feldman and members of the Dixit Lab for helpful discussions, T. Mayor and M. Petroski for reagents, D. Yansura for technical assistance, and K. Newton for editorial assistance. We acknowledge the contributions from colleagues whose work has been cited indirectly owing to space limitations. I.E.W. was supported in part by a PSTP fellowship from the University of California at Davis.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to V.M.D. (dixit@gene.com).

Structural basis for template-independent RNA polymerization

Kozo Tomita^{1,2*}, Shuya Fukai^{3*}, Ryuichiro Ishitani⁴, Takuya Ueda², Nono Takeuchi², Dmitry G. Vassilyev⁵ & Osamu Nureki^{3,6}

¹Institute for Biological Resources and Functions, National Institute of Advanced Industrial Science and Technology, 1-1-1, Higashi, Tsukuba-shi, Ibaragi 305-8666, Japan

²Department of Integrated Bioscience, Graduate School of Frontier Science, University of Tokyo, 5-1-5, Kashiwanoha, Kashiwa-shi, Chiba 277-8562, Japan

³Department of Biological Information, Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama-shi, Kanagawa 226-8501, Japan

⁴Department of Biotechnology, Graduate School of Agricultural and Life Sciences, the University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan

⁵Cellular Signaling Laboratory, RIKEN Harima Institute at SPring-8, 1-1-1 Kouto Mikazuki-cho, Sayo-gun, Hyogo 679-5148, Japan

⁶PRESTO, JST, Honcho, Kawaguchi-shi, Saitama 332-0012, Japan

* These authors contributed equally to this work.

The 3'-terminal CCA nucleotide sequence (positions 74–76) of transfer RNA is essential for amino acid attachment¹ and interaction with the ribosome^{2–4} during protein synthesis. The CCA sequence is synthesized *de novo* and/or repaired by a template-independent RNA polymerase, 'CCA-adding enzyme', using CTP and ATP as substrates⁵. Despite structural and biochemical studies^{5–8}, the mechanism by which the CCA-adding enzyme synthesizes the defined sequence without a nucleic acid template remains elusive. Here we present the crystal structure of *Aquifex aeolicus* CCA-adding enzyme, bound to a primer tRNA lacking the terminal adenosine and an incoming ATP analogue, at 2.8 Å resolution. The enzyme enfolds the acceptor T helix of the tRNA molecule. In the catalytic pocket, C75 is adjacent to ATP, and their base moieties are stacked. The complementary pocket for recognizing C74–C75 of tRNA forms a 'protein template' for the penultimate two nucleotides, mimicking the nucleotide template used by template-dependent polymerases. These results are supported by systematic analyses of mutants. Our structure represents the 'pre-insertion' stage of selecting the incoming nucleotide and provides the structural basis for the mechanism underlying template-independent RNA polymerization.

CCA-adding enzyme, also known as ATP(CTP):tRNA nucleotidyltransferase, is a unique nucleotidyltransferase^{9–11}, because it extends a specific primer (tRNA) by adding a defined sequence (CCA) in a template-independent fashion. It also monitors the status of the tRNA 3' terminus and reconstructs the CCA sequence as needed. Structural studies of CCA-adding enzyme, alone and in complex with CTP or ATP, have shown that both NTP substrates are recognized in a single catalytic pocket^{12–14}. The template-independent polymerization mechanism remains, however, to be elucidated.

On the basis of sequence homology, CCA-adding enzymes are divided into two classes, the archaeal enzyme (class I) and the eubacterial/eukaryotic enzymes (class II)¹¹. Although both classes of CCA-adding enzyme catalyse the same reaction and crossreact with each other's tRNA substrates, structural studies have shown that, other than a similar catalytic domain, the overall structures of class I and class II enzymes are fundamentally different, suggesting their divergent evolution¹³. Furthermore, in ancient and slow-evolving bacteria (*A. aeolicus*, *Deinococcus radiodurans* and *Synechocystis* sp.), the CCA-adding activity is catalysed by the two highly specific C74–C75- and A76-adding enzymes^{15,16}, suggesting that the CCA-adding activity may have been originally accomplished by two different enzymes. Phylogenetic analysis has indicated that the

modern CCA-adding enzyme may have evolved from A-adding enzymes, rather than CC-adding enzymes¹⁶.

The carboxy-terminal half of *A. aeolicus* A-adding enzyme (Aa.LC) has the same activity as the full-length A-adding enzyme both *in vitro* and *in vivo*¹⁵. The tRNA-bound Aa.LC structure resembles a seahorse, similar to the tRNA-free structures of class II CCA-adding enzymes¹², and consists of four domains referred to as the head, neck, body and tail domains^{12,17} (Fig. 1). The amino-terminal head domain is a catalytic domain that is conserved in all of the known template-independent DNA and RNA polymerases. This domain comprises antiparallel β -sheets containing the three conserved catalytic carboxylates that are thought to be essential for catalytic activity and for coordinating the Mg^{2+} ion^{9,18}. The catalytic domain of template-independent polymerases might be functionally equivalent to the palm domain of template-dependent polymerases.

One Aa.LC molecule binds one tRNA molecule. As compared with the tRNA-free structure of *Bacillus stearothermophilus* CCA-adding enzyme¹², the head and neck domains of the tRNA-bound Aa.LC adopt an expanded arrangement to accommodate the 3' end of the tRNA (Fig. 1). Aa.LC enfolds the acceptor T helix of the L-shaped tRNA molecule (Fig. 1). The body domain contacts both the T and acceptor stems, with an overall interface of 1,817 Å². Several hydrogen-bonding interactions anchor the acceptor end of the tRNA to the enzyme. The O2' atoms of the A73, G2 and C3 riboses are recognized by O δ of Asn 194, O γ of Ser 281 and N η 2 of Arg 287, respectively (data not shown). The phosphate groups of C74, C62 and C3 are recognized by N η 2 of Arg 191, N ϵ 2 of His 321 and O γ of Ser 284, respectively (data not shown).

These interactions may allow the enzyme to monitor the length of the acceptor T helix, which might explain why tRNAs (and a minihelix) are predominantly selected as primers by CCA-adding enzyme. These structural observations are consistent with the results of previous damage-selection studies using eubacterial/eukaryotic CCA-adding enzymes¹⁹ and biochemical studies using Aa.LC and a minihelix primer (data not shown). Our complex structure also supports the previous implication that the directions in which the tRNA approaches the catalytic domain are reversed between class I and class II CCA-adding enzymes^{12,13,20}.

The head and neck domains form an interdomain cleft that accommodates an incoming nucleotide and the 3' terminus of the primer tRNA. In the cleft, the electron density of the ATP analogue, AMPcPP (α,β -methyleneadenosine 5'-triphosphate), is readily

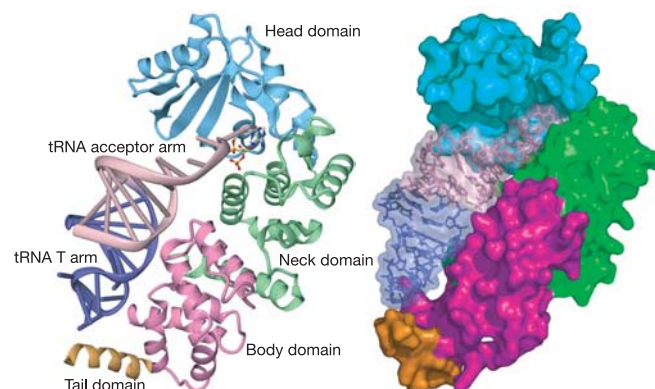


Figure 1 Overall structure of the ternary complex of Aa.LC, tRNA-CC and AMPcPP. Shown are a ribbon representation (left) and a contact surface (right). The AMPcPP molecule bound in the catalytic cleft is shown in ball-and-stick representation. tRNA is also shown in stick representation in the translucent surface on the right. The head (residues 1–137), neck (138–252), body (253–351) and tail (375–390) domains are coloured cyan, green, magenta and orange, respectively. Only the visible acceptor (pink) and T arms (blue) of the tRNA are shown.

apparent (Fig. 2a). The 2'- and 3'-OH groups of the ATP analogue hydrogen bond with the main chain CO group of Asn 109 and the main chain NH group of Gly 18, respectively (Fig. 2b). This recognition of the 2'-OH group may favour the incorporation of ribonucleotides. In the Aa.LC ternary complex, the N1 atom and the N6 amino group of the adenine base form hydrogen bonds with N η 2 of Arg 152 and O δ 2 of Asp 149, respectively (Fig. 2b). The adenine base is stabilized by stacking interactions with the side chains of Asp 105 and Arg 155, which form planar bipartite hydrogen bonds (Fig. 2c).

The recognition manner of this adenine base in the Aa.LC complex is similar to that in the *B. stearotherophilus* CCA-adding

enzyme bound to ATP¹². However, there is no hydrogen bond between Arg 152 and Glu 148 in the Aa.LC complex, whereas the corresponding hydrogen bond (between Arg 157 and Glu 153) is present in the *B. stearotherophilus* tRNA-free CCA-adding enzyme. In the tRNA-free CCA-adding enzyme, the conformational change of Glu 153 is thought to switch enzyme specificity between ATP and CTP by rotating Arg 157 (ref. 12). Therefore, the active-site structure of the present Aa.LC complex may ensure the specific incorporation of only ATP into the tRNA primer and may prevent CTP incorporation. Accordingly, the Aa.LC complex crystals dissolved when soaked in an ATP-containing solution, whereas CTP was observed in the nucleotide pocket without affecting the crystals

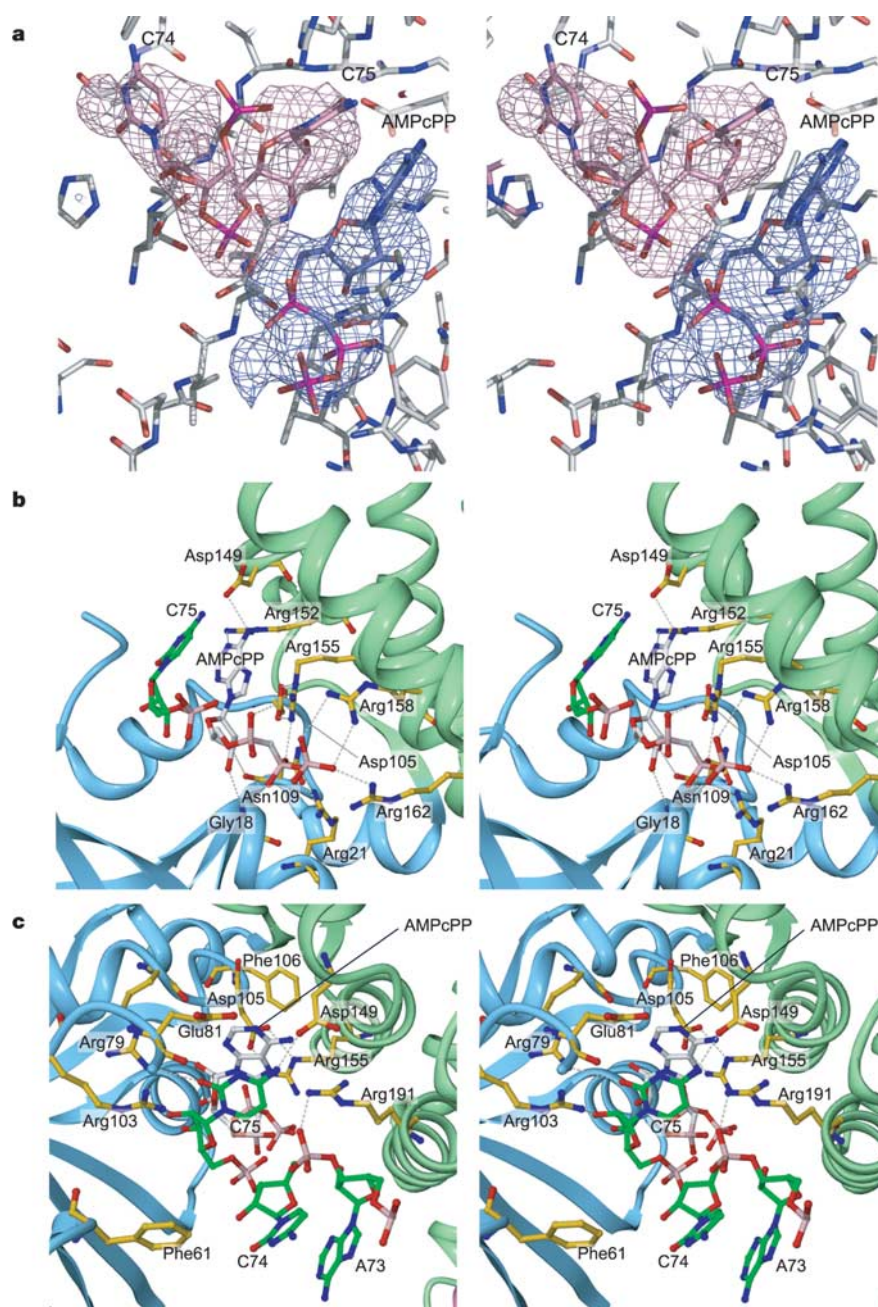


Figure 2 Stereoview of the primer C74-C75 and the incoming ATP. **a**, σ_A -Weighted simulated-annealing $F_o - F_c$ omit maps contoured at 3.5σ around C74-C75 and AMPcPP. The carbon atoms of Aa.LC, tRNA and AMPcPP are coloured white, pink and blue, respectively. **b**, Recognition of the incoming ATP. Ball-and-stick representations of tRNA C75, AMPcPP and the ATP-interacting residues are shown on the Aa.LC head and

neck domains. The colouring scheme is the same as in Fig. 1. **c**, Recognition of the C74-C75 terminus. Ball-and-stick representations of tRNA A73-C74-C75, AMPcPP and the tRNA-interacting residues are shown. Phe 106 and the Asp105-Arg155 pair, which are part of the 'stacking arc', are also shown in ball-and-stick representation. In **b** and **c**, hydrogen bonds are shown as dotted lines.

when soaked in a CTP-containing solution. This suggests that CTP, unlike ATP, does not activate Aa.LC into an 'insertion' state (see below), which may explain not only why Aa.LC does not incorporate CMP into tRNA-CC but also why Aa.LC does not have C74-C75-adding activity. In fact, CCA-adding enzymes from classes I and II add a string of CMPs in the absence of ATP⁸, whereas Aa.LC lacks this property (data not shown).

The triphosphate moiety of the incoming ATP is recognized by arginine clusters (Fig. 2b). The α -phosphate hydrogen bonds with N η 1 of Arg 155. The β -phosphate group of the ATP analogue hydrogen bonds with N η 2 of Arg 155 and N η 1 of Arg 21, whereas the γ -phosphate group hydrogen bonds with N η 1 and N η 2 of Arg 158 and N η 1 of Arg 162. An Mg²⁺ ion is not coordinated to any of the α -, β - and γ -phosphates.

On the basis of the crystal structure, we introduced a series of site-directed mutations into the catalytic pocket of Aa.LC and measured the *in vitro* AMP incorporation activity (Fig. 3). The mutagenesis results confirmed that, in addition to the putative catalytic carboxylates Asp 31, Asp 33 and Glu 74, all of the amino acids directly involved in ATP recognition are essential for incorporating AMP into tRNA lacking A76. Furthermore, Phe 106, Phe 159 and Glu 195, which interact with the ATP-recognizing residues and may form the catalytic pocket, are important for activity. We also identified amino acid residues that are necessary for AMP incorporation but are not involved in ATP recognition. These residues, Phe 61, Arg 79, Glu 81, Tyr 83, Pro 90, Arg 104 and Arg 191, are likely to recognize the C74-C75 terminus of tRNA.

The C74-C75 sequence of tRNA is a determinant for AMP incorporation into position 76 by Aa.LC: mutation of either C74 or C75 impaired the AMP incorporation rate *in vitro* (Fig. 3a). This result implies that both C74 and C75 must be recognized by Aa.LC at a specific stage during the RNA polymerization cycle. In our ternary complex structure, C74-C75 is basically flexible, in contrast to the incoming ATP analogue. The most feasible model was confirmed by the σ_A -weighted $2F_o - F_c$ and simulated-annealing $F_o - F_c$ omit maps (Fig. 2a). Furthermore, our site-directed mutagenic studies supported this atomic model of C74-C75 of tRNA (Fig. 3b). In the structure, the terminal single-stranded region, A73-C74-C75, enters the catalytic cleft, and C75 is accommodated in the complementary pocket comprising Arg 79, Glu 81, Arg 103, Arg 104, Asp 149 and Arg 191 (Fig. 2c). The O2 atom of C75 hydrogen bonds to N η 2 of Arg 79, whereas the N4 amino group

hydrogen bonds to O δ 2 of Asp 149 (Fig. 2c), which contributes to the discrimination of C75 from other nucleotides. The N η 2 atom of Arg 103 hydrogen bonds to the ribose 2'-OH group of C75 (Fig. 2c), discriminating ribose from deoxyribose at the 3'-end nucleotide of the primer, consistent with previous biochemical studies^{8,21}. Glu 81, which is also essential for AMP incorporation, is located in the proximity of the N3 atom of C75 (Fig. 2c) and may recognize C75 in the subsequent 'insertion' step (see below).

C75 is positioned adjacent to the incoming ATP analogue, and their base moieties are stacked on each other (Fig. 2). Asp 149 bridges the base moieties of the terminal C75 and the incoming ATP. This structural arrangement of the primer and the incoming nucleotide in the Aa.LC ternary complex is similar to that in the ternary complexes of rat polymerase- β and T7 RNA polymerase, which represents a snapshot of template-dependent DNA or RNA polymerization²²⁻²⁴. Notably, the side chain of Phe 106, the π -electron platform formed by Asp105-Arg155 hydrogen bonds, the adenine base of ATP, and the cytosine base of C75 form a 'stacking arc' to stabilize the location of the incoming ATP and tRNA C75 (Fig. 2c). In this context, Phe 106 and Asp105-Arg155 mimic the continuously stacked base moieties of RNA. The stacking arc leaves no room to accommodate any more incoming nucleotides, which may prevent the addition of more AMP molecules onto A76 of the tRNA.

In the Aa.LC ternary complex structure, an Mg²⁺ ion is not coordinated to either the putative catalytic carboxylates (Asp 31, Asp 33 and Glu 74) or the triphosphate moiety of the ATP analogue. The side chain of Asp 31 is not directed towards the α -phosphate group of the ATP analogue but protrudes in the opposite direction (Fig. 4a). Furthermore, the arrangement of the incoming ATP, tRNA C75 and the putative catalytic carboxylates is less compact in the Aa.LC ternary complex than in the ternary complexes of T7 RNA polymerase (Fig. 4). Therefore, the ternary complex structure may represent a state in which Aa.LC has already selected the incoming substrate ATP but is not ready to catalyse the polymerization. As described above, when ATP was soaked into the crystals instead of the non-hydrolysable AMPcPP analogue, the crystals dissolved, implying that a conformational change affecting the crystal packing had occurred. Thus, the structure should be equivalent to the 'pre-insertion' step before the reactive 'insertion' step during a single nucleotide addition cycle, as defined for template-dependent RNA polymerases^{22,23}.

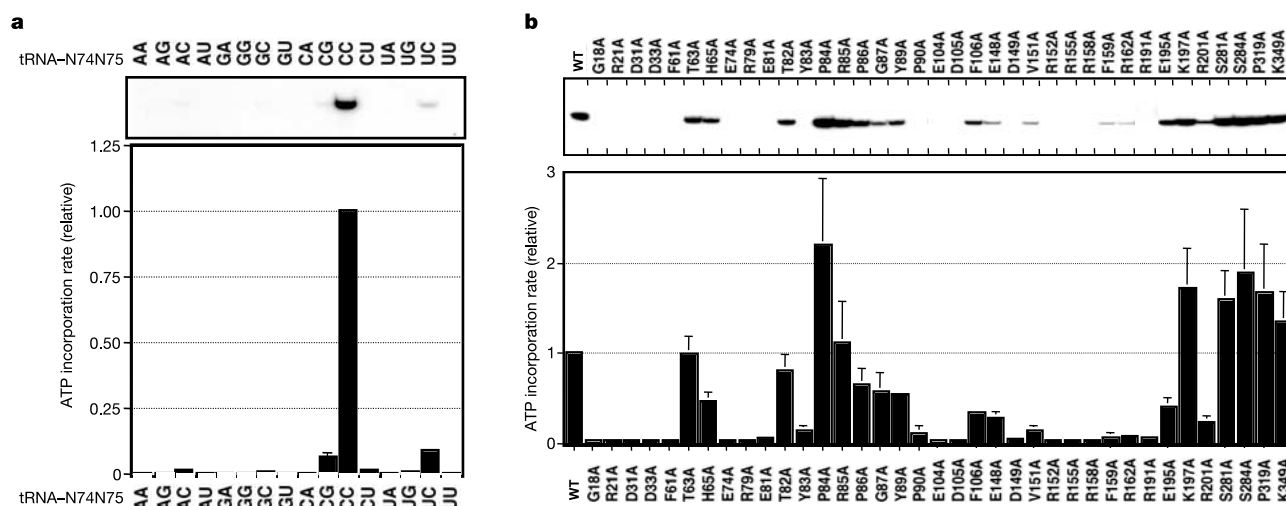


Figure 3 *In vitro* AMP incorporation. **a**, Incorporation of AMP into mutant tRNAs (tRNA-N74N75, where N is A, G, C or U). The initial rate of AMP incorporation into tRNA-N74N75 is shown. **b**, Incorporation of AMP into tRNA-CC by Aa.LC mutants. The

initial rate of AMP incorporation into tRNA-CC is shown. Error bars indicate the s.d. of three independent experiments.

In our pre-insertion complex of Aa.LC, no amino acid interacts specifically with C74, in contrast to C75. Only Phe 61, which affects the nucleotidyl-transfer activity, is located in the vicinity of the C74 base moiety (Fig. 2c), and these residues may stack together in the subsequent insertion state. In the Aa.LC complex structure, amino acid residues 83–90 are disordered. This region is close to the 3'-terminal C74–C75 sequence and is likely to be involved in the recognition of C74. Indeed, our biochemical studies showed that Tyr 83 and Pro 90 might participate in tRNA recognition (Fig. 3b). On the basis of spatial requirements, the OH group of Tyr 83 may specifically recognize the N4 or N3 atom of C74 in the subsequent insertion step. Notably, this flexible loop of the A-adding enzymes contains more proline residues than do the loops of the common CCA-adding enzymes¹⁶. This abundance of proline residues might affect the loop conformation, such that A73 or C74 of tRNA cannot enter the catalytic pocket for CMP incorporation. This may explain how the A-adding enzyme could have evolved and adapted for the more common addition of CCA.

Structural studies of T7 RNA polymerase have shown that the enzyme active site shifts between catalytically inactive 'open' and active 'closed' states during the nucleotide addition cycle. These two states of the enzyme correspond to the pre-insertion (open) and insertion (closed) binding modes for the incoming substrate and

the acceptor DNA template. The pivoting rotation of one helix (termed the 'O helix', as in the *Escherichia coli* Klenow fragment), which interacts with both the base and the triphosphate moieties of an incoming NTP at both of its ends, in the fingers domain has a key role in the conformational transition^{22,23} (Fig. 4b).

In the Aa.LC structure, the second helix in the neck domain (ATP-interacting helix, residues 151–164) is likely to be the counterpart of the O helix (Fig. 4a). This helix interacts with both the adenine and triphosphate moieties of the incoming ATP at both of its ends, as in T7 RNA polymerase. However, the directions and orientations of the helices are different and, in the Aa.LC complex, the ATP-interacting helix binds the incoming ATP more tightly and fixes the locations of both the base and triphosphate moieties (Fig. 4a). Therefore, the pivoting rotation of the ATP-interacting helix, as in T7 RNA polymerase, would not cause any productive conformational transition in the CCA-adding enzyme. Instead, in the insertion state of the CCA-adding enzyme, the head (palm) domain may move towards the active site so that the primer C74–C75 is strictly recognized and pushed towards the α -phosphate group of the incoming ATP, and the catalytic carboxylates of the head (palm) domain may interact with the triphosphate moiety through Mg^{2+} recruitment, enabling the 3'-OH group of C75 to attack the ATP α -phosphate.

The structure of a class I CCA-adding enzyme in complex with an RNA duplex mimicking the tRNA acceptor stem, and the incoming NTP, accompanies our paper²⁰. The structure represents a reactive 'insertion' step, and presents a significantly similar arrangement of C74, C75 and the incoming ATP at the catalytic site, as compared with our class II complex structure, although the methods of approach of tRNA towards the enzyme are basically different. □

Methods

Crystallization of the Aa.LC ternary complex

Aa.LC¹⁵ was overexpressed in BL21(DE3) codon plus and purified with a HiTrap chelating column, followed by anion- and cation-exchange columns. Selenomethionine-labelled Aa.LC was overexpressed in B834 (DE3) codon plus and purified as the wild-type enzyme. *Thermotoga maritima* tRNA^{Phe} lacking the terminal A76 was cloned downstream of the T7 RNA promoter and was *in vitro* transcribed using the *Bbs*I-digested plasmid. The transcript was purified with HiLoad Q-Sepharose, followed by 10% denaturing PAGE. We crystallized the complex by the hanging-drop method at room temperature in 50 mM sodium cacodylate (pH 6.0), 40 mM magnesium acetate and 30% (v/v) MPD. For the Aa.LC–tRNA–AMPcPP ternary complex, the crystals were soaked into a reservoir solution containing 2 mM AMPcPP (Sigma).

Data collection and structure determination

For phasing, a selenomethionine derivative data set was collected at the BL41XU station at SPring-8 (Harima, Japan). The crystal was cryo-protected with reservoir solution and flash-frozen in a 100-K nitrogen stream. We processed the diffraction images with the programs Denzo and Scalepack²⁵. At first, the crystal apparently belonged to the space group $P2_12_12$ with cell parameters $a = 58.79 \text{ \AA}$, $b = 109.16 \text{ \AA}$, $c = 124.90 \text{ \AA}$. The data set collected at 12.662 keV was used to locate the seven selenium atoms with the program SnB²⁶. Phase calculation was carried out with the program MLPHARE²⁷ (Supplementary Table 1). The initial phases were improved by density modification with solvent flattening and histogram matching with the program RESOLVE²⁸. The resultant electron density was readily interpretable, except for about 80 residues at the C terminus of the protein. Although the density around the tRNA was faint, the RNA double strands in the acceptor and T stems could be assigned. The model containing the protein (residues 1–81 and 92–341) and the tRNA (residues 1–7 and 49–72) was built by the program O²⁹.

Structure refinement

We subjected the initial atomic model of the complex to structure refinement using the selenomethionine derivative data set processed with $P2_12_12$. Five per cent of the total unique reflections were randomly selected for the R_{free} calculation. After three rounds of conjugate gradient minimization, the difference between the R_{work} and R_{free} values was over 10%, and the resultant $2F_o - F_c$ map was worse than the experimental map. We therefore suspected pseudo-symmetry and expanded the data from the $P2_12_12$ space group to $P2$. This assisted the refinement procedure, but the R_{free} value was still around 40%. Finally, we suspected the pseudo-merohedral perfect twinning. The introduction of the twinning parameter to the refinement markedly improved the R_{free} values and the electron density map.

The atomic model was refined with a few rounds of conjugate gradient minimization and used for molecular replacement phasing of the Aa.LC–tRNA–AMPcPP ternary complex. The native crystal of the ternary complex belonged to space group $P2$, with cell constants $a = 109.48 \text{ \AA}$, $b = 125.91 \text{ \AA}$, $c = 58.93 \text{ \AA}$, $\beta = 90.00^\circ$. Using this data set, we

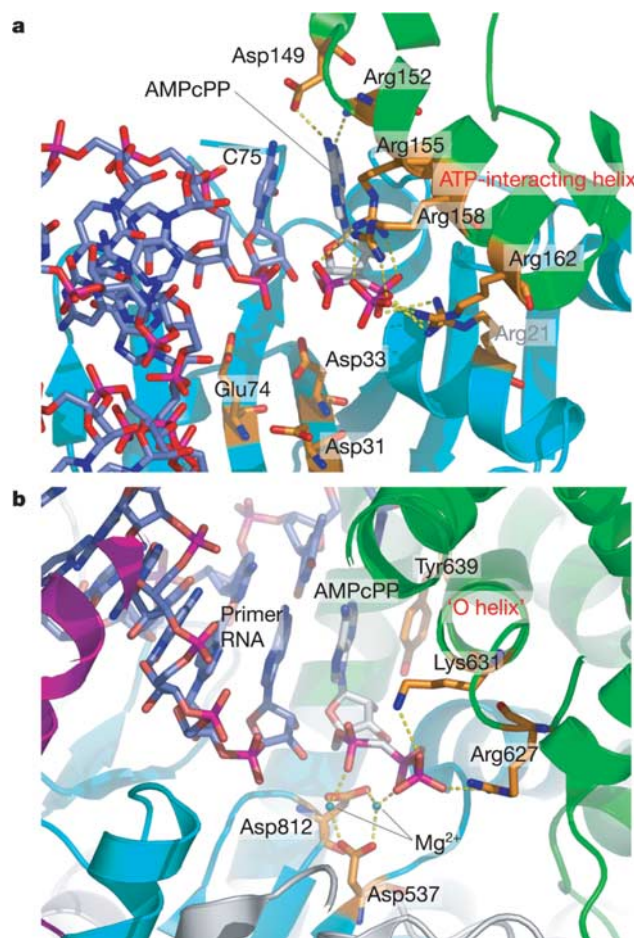


Figure 4 Comparison of template-independent and template-dependent RNA polymerases. **a**, Ball-and-stick representations of tRNA, AMPcPP, the catalytic carboxylates and the ATP-interacting residues are shown on the Aa.LC head and neck domains. The colouring scheme is the same as in Fig. 2, except that the carbon atoms of Aa.LC are coloured orange. **b**, Ball-and-stick representations of the primer RNA, the template DNA, AMPcPP, the catalytic carboxylates and the ATP-interacting residues are shown on the O helix in the T7 RNA polymerase structure. In **a** and **b**, hydrogen bonds are shown as dotted lines.

added a model of C73–C74 of the tRNA and residues 342–351 and 375–390 of Aa.LC. The refinement was carried out by conjugate gradient minimization and torsion angle dynamics simulated annealing³⁰. Iterative model building and refinement improved the model, resulting in final R_{work} and R_{free} values of 23.0% and 28.7%, respectively.

The final model includes two AMPcPP molecules and 79 water molecules. All of the refinement calculations were done with CNS³¹. Refinement statistics are shown in Supplementary Table 1. All of the figures were prepared with the programs Que (<http://www.biochem.s.u-tokyo.ac.jp/~ishitani/que/>) and PyMol (<http://pymol.sourceforge.net/>).

In vitro assays

Mutations were generated by the QuickChange mutagenesis kit (Stratagene). The mutant proteins were overexpressed and purified as described for the wild-type enzyme. We assayed activity in a solution containing 50 mM glycine (pH 8.5), 10 mM MgCl₂, 35 mM KCl, 1 mM dithiothreitol, 5% (v/v) glycerol, 0.2 μM tRNA, 10 μM ATP, 100 nM [α -³²P]ATP (3,000 Ci mol⁻¹) and 10 nM enzyme at 50 °C. After the reactions were stopped, the samples were separated by 10% denaturing PAGE. The assay conditions were developed, on the basis of the wild-type Michaelis constant (K_m) values for tRNA (2.3 μM) and ATP (47 μM), to be sensitive enough to evaluate mutant activity. Assays of mutant tRNAs (tRNA-N74N75, where N is A, G, C or U) were done as described above, except that the concentrations of tRNA-N74N75 and ATP were 5 and 250 μM, respectively.

Received 6 April; accepted 4 June 2004; doi:10.1038/nature02712.

1. Sprinzl, M. & Cramer, F. The -C-C-A end of tRNA and its role in protein biosynthesis. *Prog. Nucleic Acid Res. Mol. Biol.* **22**, 1–69 (1979).
2. Green, R. & Noller, H. F. Ribosomes and translation. *Annu. Rev. Biochem.* **66**, 679–716 (1997).
3. Kim, D. F. & Green, R. Base-pairing between 23S rRNA and tRNA in the ribosomal A site. *Mol. Cell* **4**, 859–864 (1999).
4. Nissen, P., Hansen, J., Ban, N., Moore, P. B. & Steitz, T. A. The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**, 920–930 (2000).
5. Deutscher, M. P. in *Enzymes of Nucleic Acid Synthesis and Modification*. RNA Enzymes vol. 2 (ed. Jacob, S. T.) 159–183 (CRC, Boca Raton, FL, 1983).
6. Yue, D., Weiner, A. M. & Maizels, N. The CCA-adding enzyme has a single active site. *J. Biol. Chem.* **273**, 29693–29700 (1998).
7. Shi, P. Y., Maizels, N. & Weiner, A. M. CCA addition by tRNA nucleotidyltransferase: polymerization without translocation? *EMBO J.* **17**, 3197–3206 (1998).
8. Hou, Y. M. Unusual synthesis by the *Escherichia coli* CCA-adding enzyme. *RNA* **6**, 1031–1043 (2000).
9. Holm, L. & Sander, C. DNA polymerase β belongs to an ancient nucleotidyltransferase superfamily. *Trends Biochem. Sci.* **20**, 345–347 (1995).
10. Martin, G. & Keller, W. Mutational analysis of mammalian poly(A) polymerase identifies a region for primer binding and catalytic domain, homologous to the family X polymerases, and to other nucleotidyltransferases. *EMBO J.* **15**, 2593–2603 (1996).
11. Yue, D., Maizels, N. & Weiner, A. M. CCA-adding enzymes and poly(A) polymerases are all members of the same nucleotidyltransferase superfamily: characterization of the CCA-adding enzyme from the archaeal hyperthermophile *Sulfolobus shibatae*. *RNA* **2**, 895–908 (1996).
12. Li, F. *et al.* Crystal structures of the *Bacillus stearothermophilus* CCA-adding enzyme and its complexes with ATP or CTP. *Cell* **111**, 815–824 (2002).
13. Okabe, M. *et al.* Divergent evolutions of trinucleotide polymerization revealed by an archaeal CCA-adding enzyme structure. *EMBO J.* **22**, 5918–5927 (2003).
14. Xiong, Y., Li, F., Wang, J., Weiner, A. M. & Steitz, T. A. Crystal structures of an archaeal class I CCA-adding enzyme and its nucleotide complexes. *Mol. Cell* **12**, 1165–1172 (2003).
15. Tomita, K. & Weiner, A. M. Collaboration between CC- and A-adding enzymes to build and repair the 3'-terminal CCA of tRNA in *Aquifex aeolicus*. *Science* **294**, 1334–1336 (2001).
16. Tomita, K. & Weiner, A. M. Closely related CC- and A-adding enzymes collaborate to construct and repair the 3'-terminal CCA of tRNA in *Synechocystis* sp. and *Deinococcus radiodurans*. *J. Biol. Chem.* **277**, 48192–48198 (2003).
17. Augustin, M. A. *et al.* Crystal structure of the human CCA-adding enzyme: insights into template-independent polymerization. *J. Mol. Biol.* **328**, 985–994 (2003).
18. Brautigam, C. A. & Steitz, T. A. Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes. *Curr. Opin. Struct. Biol.* **8**, 54–63 (1998).
19. Hegg, L. A. & Thurlow, D. L. Cytidines in tRNAs that are required intact by ATP/CTP:tRNA nucleotidyltransferases from *Escherichia coli* and *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **18**, 5975–5987 (1990).
20. Xiong, Y. & Steitz, T. A. Mechanism of transfer RNA maturation by CCA-adding enzyme without using an oligonucleotide template. *Nature* doi:10.1038/nature02711 (this issue).

21. Shi, P. Y., Weiner, A. M. & Maizels, N. A top-half tDNA minihelix is a good substrate for the eubacterial CCA-adding enzyme. *RNA* **4**, 276–284 (1998).
22. Temiakov, D. *et al.* Structural basis for substrate selection by T7 RNA polymerase. *Cell* **116**, 381–391 (2004).
23. Yin, Y. W. & Steitz, T. A. The structural mechanism of translocation and helicase activity in T7 RNA polymerase. *Cell* **116**, 393–404 (2004).
24. Pelletier, H., Sawaya, M. R., Kumar, A., Wilson, S. H. & Kraut, J. Structures of ternary complexes of rat DNA polymerase β , a DNA template-primer, and dCTP. *Science* **264**, 1891–1903 (1994).
25. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
26. Weeks, C. M. & Miller, R. The design and implementation of SnB version 2.0. *J. Appl. Crystallogr.* **32**, 120–124 (1999).
27. Collaborative Computational Project Number 4, The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
28. Terwilliger, T. C. Maximum-likelihood density modification. *Acta Crystallogr. D* **56**, 965–972 (2000).
29. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
30. Rice, L. M. & Brunger, A. T. Torsion angle dynamics: reduced variable conformational sampling enhances crystallographic structure refinement. *Proteins Struct. Funct. Genet.* **19**, 277–290 (1994).
31. Brunger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank A. M. Weiner for the Aa.LC plasmids; and M. Kawamoto and H. Sakai for help with data collection at SPring-8. This work was supported by Kurata Memorial Hitachi Science and Technology Foundation, Takeda Science Foundation, Foundation of Advanced Technology Institute and a Grant-in-aid for Young Scientists (to K.T.); by Asahi Glass Foundation (to S.F.); by a grant from the Ministry of Education, Culture, Sports, Science and Technology (to N.T.); and by a PRESTO Program grant from Japan Science and Technology and a Naito Foundation grant (to O.N.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to O.N. (onureki@bio.titech.ac.jp). Atomic coordinates have been deposited in the Protein Data Bank under accession number 1VFG.

erratum

Single-crystal metallic nanowires and metal/semiconductor nanowire heterostructures

Yue Wu, Jie Xiang, Chen Yang, Wei Lu & Charles M. Lieber

Nature **430**, 61–65 (2004).

In Fig. 2b of this Letter, the units of the abscissa should be volts (V). In Fig. 3a, the step labels (3) and (4) should be interchanged. □

Contacts

Publisher: Ben Crowe
Editor: Paul Smaglik
Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Naturejobs Sales Director:
Nevin Bayouni (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)
Andy Douglas (4975)
Frank Phelan (4944)

Scandinavia/ Spain/ Portugal:

Evelina Rubio Håkansson (4973)

Natureevents: Silke Opstrup (4994)

France/ Switzerland:

Amelie Pequignot (4974)

Production Manager:

Billie Franklin
To send materials use London
address above.
Tel +44 (0) 20 7843 4814
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Stefan Hales

European Satellite Office

Germany/ Austria/ Italy/

The Netherlands/ Belgium:

Patrick Phelan, Odo Wulffien
Tel + 49 89 54 90 57 11/-2
Fax + 49 89 54 90 57 20
e-mail: p.phelan@nature.com
o.wulffien@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager:

Peter Bless

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Haraikatamachi,
Shinjuku-ku,
Tokyo 162-0841
Tel +81 3 3267 8751
Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Rinoko Asami
e-mail: rasami@naturejp.com

naturejobs

The accidental science writer

When I returned to my graduate-school stomping grounds at the University of Wisconsin at Madison, last month, I was reminded of what I now recognize was a critical career juncture. University officials and biotech executives asked me about my connections to the Badger State — and how I reached my present position.

My stock answer for such questions always begins: “I got hit by a nun.” While eking out an existence at a small Wisconsin daily newspaper in the late 1980s, my career was in danger of stalling. Newspapers in the United States were merging and down-sizing, and no one seemed to be hiring. A mentor from my undergraduate days suggested graduate school and, although I didn’t exactly relish taking a few more years out of my working life, I applied.

But my decision to attend was made for me when, on the way to my normal reporting duties, a nun smashed into the rear of my car as I was signalling to turn into the county courthouse. The impact hurled me and my car into the other lane of traffic, where I got smacked again, writing off the vehicle.

A few weeks after I walked away from the accident, I was accepted into the University of Wisconsin’s mass-communication programme. A small cheque from the insurance company financed my move, and a part-time job as a teaching assistant provided some income. As well as experience in front of a class, my spell at university allowed me to learn about biomedical research, ethics and history; write about intellectual property and technology transfer; and publish my own research on interactive health communication.

All of those experiences gave me skills and knowledge that at the time I never thought I’d use again, but which I now draw on nearly every day. And the circumstances taught me to take advantage of opportunity — even when it hits you, unawares, from behind.

Paul Smaglik
Naturejobs editor



Contents

SPECIAL REPORT

Back to basics: from industry to academia **p706**

WWW.NATUREJOBS.COM

Career centre
Information on the scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS

Scientists are returning from industry to academia with an eye on efficiency and applied science.

Kendall Powell explores the attitude adjustments.



Back to basics: Jeffrey Peng returned to academia after a decade in industry.

In the early 1990s, as Jeffrey Peng was completing a postdoc in nuclear magnetic resonance (NMR) chemistry, he decided that “academic funding stank”. So, when a young but exceptional company, Vertex Pharmaceuticals in Cambridge, Massachusetts, advertised a physical NMR position, Peng jumped at the chance. Even so, he knew his scientific interests leaned heavily towards the theoretical. Now, ten years later, he says he has finally given in to chasing his scientific demons and returned to academia as an assistant professor in the department of chemistry and biochemistry at the University of Notre Dame in Indiana.

The return to academia was hard, says Peng, but necessary because he wanted to pursue his own research. But the commercial world taught him skills that may make his academic life easier — especially in terms of efficiency and broader scientific thinking.

Peng represents a class of returners exploring basic science questions that they couldn’t forget during their time in industry. From their time in applied, product-driven science, they bring new attitudes and approaches to the academic world of hypothesis-driven research and scientific training. Others have brought an industry perspective to posts at the head of departments, schools and institutes.

These returners bring fresh ideas to running individual research groups and teaching as well. Daniel Sem was vice-president for biophysics at Triad

Therapeutics in San Diego before taking a position at Marquette University in Milwaukee, Wisconsin, to study a set of liver enzymes called cytochrome P450s. These enzymes can metabolize certain potential drugs before they have a chance to act in the body.

Working at Triad showed him how industry focuses on predicting which drugs will be metabolized, rather than on how the enzymes work. Now, he says, he is straddling the two worlds, trying to move his basic-science questions in the direction of applied science because he knows what direction to take. The ability to straddle academia and industry is perhaps most important in chemistry, as technology and philosophy differ drastically within this discipline between sectors (see *Nature Rev. Drug Disc.* 3, 719; 2004).

FRESH PERSPECTIVES

Jamie Williamson, a biochemist at the Scripps Research Institute in La Jolla, learned how to think in terms of project management while trying to start up his own company. “Figure out what the person across the table wants and needs,” he says. It may sound simple, he adds, but it holds true for business deals and scientific collaborations.

Lessons like that are now being taught at the highest levels as several industry leaders have recently returned to academic posts. One such mover, Roy Vagelos, came full-circle. He once instilled academic

Myth busters

1 You’ll never publish anything (good) in industry

“At younger biotech companies, there is an early phase where you can publish a lot that is competitive. After a company becomes successful, however, information can come in but should not go out.”

— Jeffrey Peng, Vertex and the University of Notre Dame.

2 You can’t be creative in industry

“You have to think creatively within a different framework. My lab has graduated people working away in the San Diego biotech industry. They are happy, challenged, fulfilled, interested and excited by what they do.”

— Jamie Williamson, the Scripps Research Institute.

3 Once you leave, you can never come back to academia

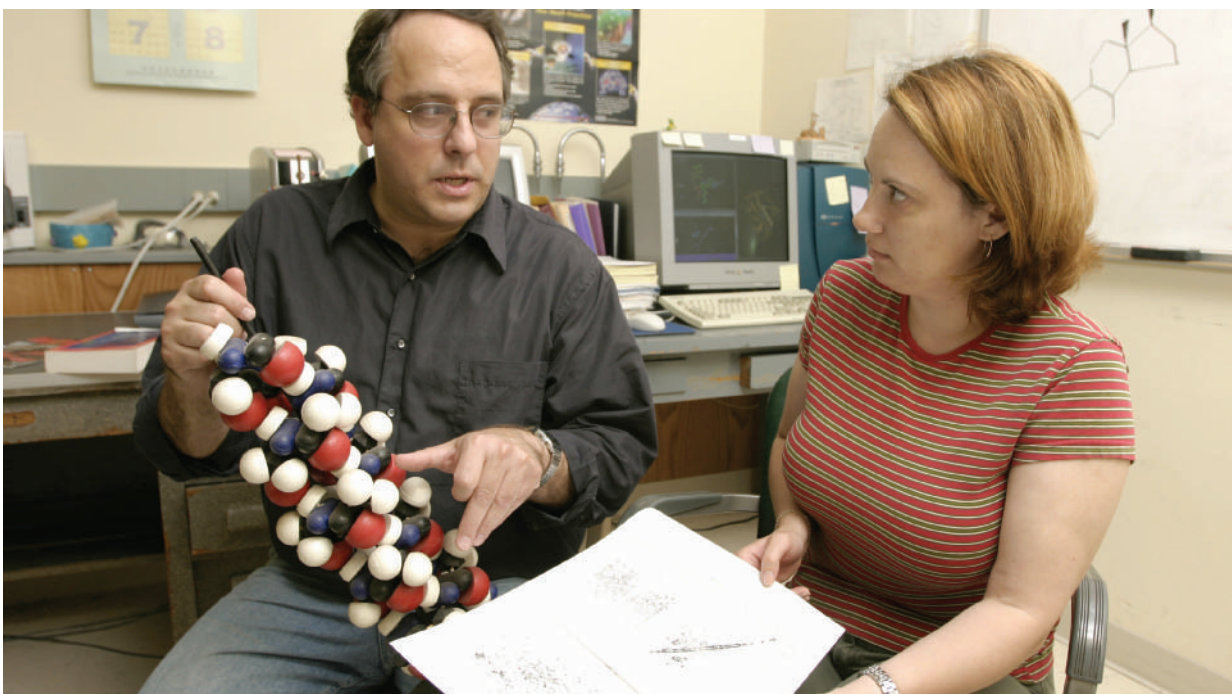
“Don’t lose touch with the people still in academia. You have to keep up with the science and resist the temptation to just watch the stock. You have to have an intrinsic interest in the science and work harder to keep it up.”

— Jeffrey Peng, Vertex and the University of Notre Dame.

4 Moving to industry means you couldn’t cut it in academia

“The research I did in a company was far broader and had far more impact on people’s lives than anything I could have done in even a very productive academic lab. Being able to interface with patients drove me forward.”

— Gail Naughton, ATS and San Diego State University.



Applied knowledge: Daniel Sem has found that his industrial experience has changed the way he teaches and motivates his students.

values at Merck in Whitehouse Station, New Jersey. Now he is working in the opposite direction; since retiring as Merck's chief executive in 1994, Vagelos has advised universities on how to integrate graduate programmes from different departments or campuses so that faculty members and students can work together industry-style.

"Identify the right people, put it in their hands, stand back and support them like hell," says Vagelos. "That is a great industrial perspective that reaches into both research and education." This key guiding principle from his years at Merck has informed his efforts at the University of Pennsylvania, Rutgers University and, currently, Columbia University Medical Center.

Gail Naughton, former co-founder and president of Advanced Tissue Sciences (ATS) — once a leading company in tissue engineering — has brought her business acumen back to an altogether different academic post. In 2002, Naughton became the dean of the College of Business Administration at San Diego State University. Her background was a match for the programme situated in the high-tech industry hub.

COMBINED EFFORT

At ATS, Naughton recognized that scientists needed substantial on-the-job training in business skills. Few science programmes even mention concepts such as quality control, regulatory affairs or intellectual-property protection, although these are critical to industry. "There's a tremendous opportunity for academics to partner with the life-science industry and train people much more specifically," says Naughton.

Her tenure as dean has moved the university in that direction with a combined MBA–PhD programme, which requires a one-year industry internship, a PhD thesis on a product or platform, and a business plan to carry the discovery to market as the MBA component.

Richard Sykes, too, is using his industrial experience to reshape a university. The former Glaxo Wellcome chairman, now rector of Imperial College

London, has encouraged technology transfer and is trying to create an entrepreneurial culture among students and faculty members alike. He even envisages creating an MSc in business administration to help researchers in all sectors become better managers (see www.nature.com/naturejobs/channels/graduate).

Perhaps the greatest benefit from the cross-pollination, though, is exposing students and postdocs to industry, whether through lectures by company leaders, courses, internships or collaborations, says Vagelos. That should demystify industry and reveal its career potential (see 'Myth busters', opposite).

Graduate students need exposure to industrial labs to see that they are "full of normal human beings" doing a range of exciting activities, says Simon Bright, director of the University of Warwick's department of horticultural science, Warwick HRI, in Coventry, UK. He took the post after 18 years at the agricultural biotech company Syngenta. "I want to encourage the group to be a little bit brave because there are some rewards to being brave," he says. He sees the next generation of scientists as having a more entrepreneurial spirit.

Naughton says that students considering a career in industry should seek out PhDs who have been successful there, to "have a basic, blunt discussion on the challenges and skill sets of industry". She also recommends doing a one-year postdoc in industry to get a deep sense of the business while still publishing and building your curriculum vitae.

Peng would like to impart to students the wisdom he gained from his ten years at Vertex. "If you want to contribute to the huge achievement of putting a drug out, there will be better science and infrastructure in a company," he says. "But if you've got some scientific question you've always wanted to answer, you will ultimately satisfy that urge in academia." That kind of advice will better prepare the next generation of scientists to work in a world where the fence between academia and industry no longer exists. ■

Kendall Powell is freelance science writer based in Broomfield, Colorado.